

WestminsterResearch

<http://www.westminster.ac.uk/westminsterresearch>

**Automatic Pancreas Segmentation and 3D Reconstruction for
Morphological Feature Extraction in Medical Image Analysis
Asaturyan, H.**

This is an electronic version of a PhD thesis awarded by the University of Westminster.

© Ms Hykoush Asaturyan, 2021.

The WestminsterResearch online digital archive at the University of Westminster aims to make the research output of the University available to a wider audience. Copyright and Moral Rights remain with the authors and/or copyright owners.

Automatic Pancreas Segmentation and 3D Reconstruction for Morphological Feature Extraction in Medical Image Analysis

Hykoush Asaturyan

A thesis submitted in partial fulfilment of the requirements
of the University of Westminster
for the degree of Doctor of Philosophy

School of Computer Science and Engineering
College of Design, Creative and Digital Industries
University of Westminster, London, UK

This research programme was completed within the
Computational Vision and Imaging Technology Research Group.

1st March 2021

ABSTRACT

The development of highly accurate, quantitative automatic medical image segmentation techniques, in comparison to manual techniques, remains a constant challenge for medical image analysis. In particular, segmenting the pancreas from an abdominal scan presents additional difficulties: this particular organ has very high anatomical variability, and a full inspection is problematic due to the location of the pancreas behind the stomach. Therefore, accurate, automatic pancreas segmentation can consequently yield quantitative morphological measures such as volume and curvature, supporting biomedical research to establish the severity and progression of a condition, such as type 2 diabetes mellitus. Furthermore, it can also guide subject stratification after diagnosis or before clinical trials, and help shed additional light on detecting early signs of pancreatic cancer. This PhD thesis delivers a novel approach for automatic, accurate quantitative pancreas segmentation in mostly but not exclusively Magnetic Resonance Imaging (MRI), by harnessing the advantages of machine learning and classical image processing in computer vision. The proposed approach is evaluated on two MRI datasets containing 216 and 132 image volumes, achieving a mean Dice similarity coefficient (DSC) of $84.1 \pm 4.6\%$ and $85.7 \pm 2.3\%$ respectively. In order to demonstrate the universality of the approach, a dataset containing 82 Computer Tomography (CT) image volumes is also evaluated and achieves mean DSC of $83.1 \pm 5.3\%$. The proposed approach delivers a contribution to computer science (computer vision) in medical image analysis, reporting better quantitative pancreas segmentation results in comparison to other state-of-the-art techniques, and also captures detailed pancreas boundaries as verified by two independent experts in radiology and radiography. The contributions' impact can support the usage of computational methods in biomedical research with a clinical translation; for example, the pancreas volume provides a prognostic biomarker about the severity of type 2 diabetes mellitus. Furthermore, a generalisation of the proposed segmentation approach successfully extends to other anatomical structures, including the kidneys, liver and iliopsoas muscles using different MRI sequences. Thus, the proposed approach can incorporate into the development of a computational tool to support radiological interpretations of MRI scans obtained using different sequences by providing a “second opinion”, help reduce possible misdiagnosis, and consequently, provide enhanced guidance towards targeted treatment planning.

Acknowledgments

I would like to express my heartfelt gratitude and deep appreciation to my Director of Studies, Dr Barbara Villarini, who has offered continuous guidance, support and dedicated supervision throughout the development of my PhD project detailed in this thesis.

Moreover, I would like to sincerely thank my other supervisors, Professor Sophie Triantaphillidou, Dr Alexandra Psarrou and Professor Jimmy D. Bell, for their guidance throughout my PhD project development and their unique perspectives and support towards finalising this thesis.

I would also like to express my gratitude to Dr Sila Kurugol for her supervision during my position at Boston Children's Hospital in affiliation with Harvard Medical School.

To end, I wish to thank Professor E. Louise Thomas, Dr Julie A. Fitzpatrick and Professor Antonio Gligorievski for their expert qualitative feedback.

Author's Declaration

I declare that all the material contained in this thesis is my own work, and has not been submitted, in whole or in part, for any other degree or professional qualification. The source of images provided in other publications has been appropriately referenced in this thesis.

By: Hykoush Asaturyan

Table of Contents

Abstract	2
Acknowledgments	3
Author's Declaration	4
Table of Contents	9
List of Tables	10
List of Figures	19
Glossary	21
1 Introduction	22
1.1 Introduction	22
1.2 Objective of Research	24
1.3 Introducing Original Contributions to Knowledge	25
1.4 List of Publications	26
1.5 Organisation of Remaining Chapters	28
1.6 Conclusion	29
2 Background and Rationale	30
2.1 Introduction	30
2.2 Significance of Pancreas Segmentation in CADx	31
2.2.1 Monitoring Diabetes	32
2.2.2 Pancreatic Cancer and Pancreatitis	35
2.3 Challenges	37
2.3.1 Structural Variability	38
2.3.2 MRI vs CT Modality	39

2.3.3	Dataset Limitations	40
2.4	Conclusion	41
3	Related Work and Contributions	42
3.1	Introduction	42
3.1.1	Quantitative Measure of Accuracy	42
3.2	State-of-the-art Methods	43
3.2.1	Multi-atlas Label Propagation and Statistical Shape Modelling	44
3.2.2	Integrating Convolutional Neural Networks	46
3.2.3	Baseline Architectures: U-Net, DenseNets and Residual	47
3.3	Deep Learning for Medical Image Segmentation	51
3.3.1	DenseVNet	52
3.3.2	Recurrent Saliency Transformation Network	54
3.3.3	Multi-atlas Combined FCN Approach	55
3.3.4	Optimising the Jaccard Index Directly	56
3.3.5	Limitations	58
3.4	Original Contributions to Knowledge	58
3.4.1	(a) Digital Contrast Enhancement	58
3.4.2	(b) Hybrid Energy-minimising Segmentation	59
3.4.3	(c) Multi-level Geometrical Descriptor Analysis and Tissue Classification	59
3.4.4	(d) Loss Function	60
3.4.5	Attributes of Contributions	60
3.5	Conclusion	61
4	Method 1: Morphological and Multi-level Geometrical Descriptor Analysis	62
4.1	Introduction	62
4.2	Tissue Enhancement and Elimination	64
4.2.1	Digital Contrast Enhancement	64
4.2.2	Identify Major Region of Interest	65
4.3	Rough Segmentation of Pancreas	66
4.3.1	Edge Detection and Boundary Matching	70
4.4	Refinement	73

4.4.1	Level 1: Morphological Operations on Distinct Contours	75
4.4.2	Level 2: Localisation and Positioning of Contours	76
4.4.3	Level 3: Centre Landmarks of Distinct Contours	76
4.5	Conclusion	80
5	Method 2: Integrating Deep Learning	81
5.1	Introduction	81
5.2	Training Phase	82
5.2.1	Random Forests: Identify Major Pancreas Region	84
5.2.2	Detect Targeted Pancreas Region	84
5.3	Testing Phase	91
5.3.1	Predict Main Pancreas Region	91
5.3.2	3D Segmentation and Refinement	91
5.4	Transitioning from Deep Learning in 2D to 3D	92
5.4.1	Training Stage	95
5.4.2	Testing Stage	96
5.5	Conclusion	96
6	Experimental Tools, Results and Analysis	98
6.1	Introduction	98
6.2	Datasets	99
6.2.1	MRI	99
6.2.2	CT	99
6.3	Technical Tools, Libraries and Frameworks	99
6.3.1	MATLAB	100
6.3.2	TensorFlow and Keras	102
6.4	Evaluation	102
6.4.1	Method 1: Morphological and Multi-level Geometrical Descriptor Analysis	103
6.4.2	Method 2: Integrating Deep Learning	103
6.4.3	Comparison between Method 1 and 2	105
6.4.4	3D-Method 2	113
6.4.5	Comparison between Method 2 and 3D-Method 2	113
6.5	Conclusion	115

7	Discussion and Improvements	117
7.1	Introduction	117
7.2	Discussion	117
7.2.1	Quantitative Assessment	118
7.2.2	Qualitative Assessment	123
7.2.3	Computational Efficiency	124
7.3	Potential Optimisation	125
7.3.1	Improvements to Architecture in Method 2	126
7.4	Conclusion	127
8	Generalisation and Future Work	128
8.1	Introduction	128
8.2	Impact of Extracting Morphological Features	129
8.3	Proposed Generalised Segmentation Framework	130
8.3.1	Dataset	131
8.4	Liver	131
8.4.1	Experimental Results and Analysis	132
8.5	Iliopsoas Muscles	134
8.5.1	Experimental Results and Analysis	134
8.6	Kidneys	136
8.6.1	Dataset	136
8.6.2	Experimental Results and Analysis	138
8.7	Future Work	141
8.8	Conclusion	141
9	Conclusions	143
9.1	Introduction	143
9.2	Challenges and Motivation	144
9.2.1	Comparison to Related Work in Literature	144
9.3	Impact of Original Contributions	146
9.4	Comparison Between Different Proposed Methods	147
9.5	Employing Contributions into Future Work	148
9.5.1	Extension to Multiple Modalities	148
9.6	Conclusion	149

Appendix A: Exemplars of CT and MRI Slices from Distinct Image Volumes	153
Appendix B: Deep Learning Network Development and Convolutional Neural Network Architecture	156
Appendix C: Qualitative Expert Review of Pancreas Contouring Outcomes	168
Appendix D: Quantitative Morphological Features of Pancreas Volume and Curvature	170
Appendix E: Number of Combinations for Contours with Non-infinity Gradient	172
Appendix F: Example of Level 3 Refinement	176
Appendix G: Derivation of Modified Hausdorff Loss	180
References	203

List of Tables

6.1	Comparative summary of DSC, JI, PC, RC and Hausdorff (Haus.) results shown as mean \pm standard deviation [lowest, highest] in MRI-A, MRI-B and CT-NIH using Methods 1, 2 and 3D-2.	107
7.1	Overall DSC, JI, PC and RC shown as mean \pm standard deviation [lowest, highest] for automatic pancreas segmentation methods in CT modality image volumes. The value of N represents the dataset size.	119
7.2	Overall DSC, JI, PC and RC shown as mean \pm standard deviation [lowest, highest] for automatic pancreas segmentation methods in MRI modality image volumes. The value of N represents the dataset size.	120
7.3	Overall DSC and JI are shown as mean \pm standard deviation (%) for automatic pancreas segmentation methods using the publicly available CT dataset	121
7.4	Overall DSC and JI are shown as mean \pm standard deviation (%) for automatic pancreas segmentation methods using the MRI-A dataset	121
7.5	Overall DSC and JI are shown as mean \pm standard deviation (%) for automatic pancreas segmentation methods using the MRI-B dataset	122

List of Figures

1.1	Overview of sequence of steps towards automatic pancreas segmentation in Magnetic resonance imaging (MRI) volumes, and 3D rotational reconstruction of the segmented organ: (a) Sequence of original 2D axial image ‘slices’ from an MRI volume; (b) Sequence of segmented (red) 2D image ‘slices’ from the MRI volume; (c) 360-degree 3D digital reconstruction visualisation of the segmented pancreas.	23
2.1	A pancreatic cyst, marked by an arrow, as shown on an MRI. Only a small percentage of such cysts lead to cancer [188].	33
2.2	MRI scans of tumour lesions. A (T2 axial view) and B (T2 coronal view) highlight a primary tumour lesion 2 months before the initial diagnosis, which identifies a cystic lesion at the location between body and tail of the pancreas with a size of 43 mm \times 38 mm \times 30 mm (A) and dilatation of the pancreatic duct of 13 mm (B). C (T1 axial) and D (T1 coronal) highlight a relapsed hepatic lesion 20 months after an after initial diagnosis [189].	34
2.3	Pancreatic pseudocyst (Ps) in a man in his early 40s known with gallstones [190, 193]. (a) CT axial (b) T1-weighted (c) and T2-weighted MRI, obtained 6 weeks after the beginning of an acute pancreatitis condition, highlights a homogenous-kind fluid collection (Ps) with a fibrous capsule pancreas (P).	36
2.4	Location of pancreas in abdominal cavity [158]. The pancreas lies on the posterior abdominal wall. Notice that the pancreas is located just behind the stomach in the back of the upper left abdomen. The head of the pancreas lies within the duodenum curve, which is the first section of the small intestine. The tip of the pancreas extends across the abdominal cavity.	38

2.5	The expert-led manual (ground-truth) segmentation of the pancreas is contoured in red. Column (a) displays an axial slice in a CT scan. Columns (b) and (c) display two axial slices from two different MRI scans. Notice the finer image quality of CT in comparison to MRI that is significantly coarse and suffers from greater blurred boundaries between the pancreas, and duodenum.	40
3.1	Three different segmentation outcomes (green) overlapping the same ground-truth (red) representing a 3D pancreas reconstructed model. Column (a) suffers from excess false-positive segmentation labels; column (b) suffers from false-negative labels; column (c) has significantly less <i>both</i> false-positive and false-negative labels segmentation labels than (a) and (b).	43
3.2	Anatomical relationships of the pancreas with surrounding organs and structures. The head of the pancreas lies in the loop of the duodenum as it leaves the stomach; the tail of the pancreas lies near the spleen; and the body of the pancreas lies posterior (nearer the rear) to the outer region of the stomach between the tail [176].	45
3.3	U-Net architecture as originally described in [171]. Every blue block represents a multi-channel feature map, in which the number of channels is denoted on top of the block. All white blocks represent copied feature maps and the arrows represent the different operations.	48
3.4	DenseNets architectures for image classification of ImageNet [180]. The growth rate for all the networks is $k = 32$. Each “conv” layer corresponds the sequence of layers: Batch Normalisation, rectified linear units (ReLU) and convolution.	49
3.5	The residual architecture implemented for ImageNet [178]. From left to right, the VGG-19 model [97] serves as a reference; a plain network with 34 parameter layers; and a residual network with 34 parameter layers. The dotted shortcuts represent an increase in the dimensions from input to output.	50
3.6	A building block for residual learning. Whereas in traditional neural networks, each layer feeds into the next layer, each layer in a network with residual blocks feeds into the next layer and directly into the layers about some “jumps” away [178].	51
4.1	Overview of methodology: three main stages are proposed for automatic pancreas segmentation in abdominal CT and MRI scans (volumes). . .	63

4.2	Visualisation of results for a slice from image volume. (a) Original MRI slice, (b) MRI slice after contrast enhancement, (c) Closed red outline encapsulates major pancreas region, (d) Segmentation following max-flow and min-cuts approach, (e) Boundary detection using structured forest learning, (f) Final contour segmentation after refinement.	64
4.3	Visualisation of three levels for fine pancreas extraction in three different slices from three different image volumes. Column (a) displays the original slice after initial segmentation. Column(b) displays the resultant slice after first level of refinement, now contained in a bounding box against a “trail map”. Column (c) displays the resultant slice after second level of refinement, now contained in a bounding box that identifies gradient between combinational distinct contours. Column (d) displays final, resultant slice after third level of refinement.	74
4.4	Visualisation of three refinement levels in three different slices (in three different image volumes). Top row and bottom row display distinct contours before and after processing. Column (a) displays first level of morphological operations on distinct contours, highlighting measurements of area, spatial aspect ratio and triangularity. Column (b) displays second level of positioning contours on “trail map” within bounding box containing all contours. Column (c) displays combinational connectivity between centre landmark points and respective gradients between all distinct contours.	74
4.5	Column (a) provides a visual example of a slice from an image volume in a training dataset. Computations include the gradient and distance between combinational area centres and the respective ratios of combinational areas. Column (b) provides a summarised visual representation of “ground-truth characteristics” generated for sets R_1 , R_2 , R_3 and R_4	78
4.6	Visual example of three contour combination, BCD , which represents the 7th in 10 different combinations from five distinct contours. Three out of five contours are presumed valid, i.e. “pancreas” and the other two otherwise.	79

5.1	Overview of Method 2. Top diagram highlights the training phase consisting of two main parts, (A) train a random forest to identify major pancreas regions in all training image volumes; (B) train a deep learning model to recognise pancreatic features within an image volume, and later perform pixel-wise classification on a test image volume as “pancreas” or “non-pancreas”. The bottom image highlights the testing phase and consists of three main stages, (1) detect targeted pancreas region using the trained deep learning model using pixel-wise classification; (2) perform 3D segmentation of predicted pancreas region to extract distinct contours that may lie near but are not part of the pancreas; (3) perform refinement to identify non-pancreatic contours for removal from final segmented pancreas volume.	83
5.2	Overview of the deep learning model that has been incorporated into Method 2. An encoder downsamples the CT/MRI input through convolution, batch normalisation and ReLU, producing feature maps that represent unique pancreatic features (e.g. texture, boundaries, etc). A decoder upsamples its input using the transferred pooling indices from its corresponding encoder to generate sparse feature maps. From here, convolution is performed with a trainable filter of weights to density the feature map. The resulting decoder output feature maps are fed to a soft-max classifier for 2-channel pixel-wise classification of the input image as “pancreas “ or “non-pancreas”.	85
5.3	A 2×2 max-pooling is applied in the Encoder stage. The maximum value and pooling index inside this 2×2 region is retained and propagated to the next layer. In upsampling, in the Decoder stage, the max-pooled 1×1 feature is placed into to the exact location of its corresponding pooling index.	85
5.4	Detailed overview of the deep learning architecture. Each round rectangle is a layer, with the number on the right side indicating the number of channels (or feature maps in the case of convolutional layers). Each convolution (<i>Conv</i>) layer contains a set of 3×3 filters, stride 1, padding 1. Each max-pooling (<i>Pooling</i>) layer uses 2×2 window with stride 2. <i>Batch N. + ReLu</i> refers to a batch normalization layer followed by ReLu activation. During max-pooling, the index of the maximum feature in each 2×2 area is saved. For each <i>Upsampling</i> layer, the 1×1 feature is placed in the exact location where the corresponding max-pooling index is located. The last layer is a soft-max classifier with 2-channels, one relating to “pancreas” and the other “non-pancreas”. The predicted segmentation relates to the class with maximum probability at each pixel.	86

5.5	Method 2 testing phase. Column (a) displays four different slices in four different and distinct image volumes that are fed into the trained deep learning model. Column (b) reveals the predicted pancreas region mask for each corresponding input slice. Column (c) achieves max-flow and min-cuts segmentation for each input slice. Column (d) displays the pancreas region after the mask and post-processing has been applied. Column (e) reveals the segmentation contouring (green) against the ground-truth (red).	92
5.6	Overview of 3D-Method 2. The training stage simultaneously develops a network (3D Rb-UNet) for localising the pancreas, and a segmentation network (3D Tiramisu) to predict the labels that correspond to “pancreas” and “non-pancreas” tissue. The testing stage processes an original 3D or 4D image volume, performs a coarse segmentation to generate a bounding box capturing the main pancreas region and then processes the cropped image volume to predict the labels of that organ.	93
5.7	Tiramisu architecture blocks of fully convolutional DenseNets include (from left to right) the layer used in the model, Transition Down (TD) and Transition Up (TU) [179].	94
5.8	Architecture details of the Tiramisu (FC-DenseNet103) model, which is built from 103 convolutional layers [179]. The notations are as follows: DB stands for Dense Block, TD stands for Transition Down, TU stands for Transition Up, BN stands for Batch Normalisation and m corresponds to the total number of feature maps at the end of a block. c stands for the number of classes.	94
6.1	Overview of MRI segmentation accuracy results (in DSC) of models that are trained with different loss functions. The proposed Hausdorff-Sine Loss performs better across thresholds in range [0.05, 0.95] in comparison to the conventional cross-entropy and other loss functions.	104
6.2	Overview of CT-NIH segmentation accuracy results (in DSC) of models that are trained with different loss functions. The proposed Hausdorff-Sine Loss performs better across thresholds in range [0.05, 0.95] in comparison to the conventional cross-entropy and other loss functions.	104
6.3	ROC curves averaged on MRI-A and MRI-B using the deep learning model in Method 2 trained via the proposed Hausdorff-Sine loss, and other loss functions including Hausdorff (alone), Cross-entropy, Dice and Jaccard.	106

6.4	ROC curves averaged on CT-NIH using the deep learning model in Method 2 trained via the proposed Hausdorff-Sine loss, and other loss functions including Hausdorff (alone), Cross-entropy, Dice and Jaccard. . .	106
6.5	Segmentation results in four different CT image scans (volumes) using Method 1. Every column corresponds to a single CT volume. From left, first row displays sample CT axial slices with segmentation outcome (green) against ground-truth (red), and computed DSC; second row displays 3D reconstruction of entire pancreas (green) segmentation against its ground-truth (red) with computed DSC.	109
6.6	Segmentation results in four different CT image scans (volumes) using Method 2. Every column corresponds to a single CT volume. From left, first row displays sample CT axial slices with segmentation outcome (green) against ground-truth (red), and computed DSC; second row displays 3D reconstruction of entire pancreas (green) segmentation against its ground-truth (red) with computed DSC.	109
6.7	Segmentation results in six different MRI scans (volumes) for Method 1 (Morphological and Multi-level Geometrical Descriptor Analysis). Every column corresponds to a single MRI volume. From left, first row displays sample MRI axial slices with segmentation outcome (green) against ground-truth (red), and computed DSC; second row displays 3D reconstruction of entire pancreas (green) segmentation against its ground-truth (red) with computed DSC.	110
6.8	Segmentation results in six different MRI scans (volumes) for Method 2 (Deep Fusion). Every column corresponds to a single MRI volume. From left, first row displays sample MRI axial slices with segmentation outcome (green) against ground-truth (red), and computed DSC; second row displays 3D reconstruction of entire pancreas (green) segmentation against its ground-truth (red) with computed DSC.	111
6.9	Dice similarity coefficient (DSC) and Jaccard index (JI) box plots for datasets CT-NIH, MRI-A and MRI-B using Method 1.	111
6.10	Dice similarity coefficient (DSC) and Jaccard index (JI) box plots for datasets CT-NIH, MRI-A and MRI-B using Method 2.	112
6.11	Recall against precision plots for datasets CT-NIH, MRI-A and MRI-B using Method 1.	112
6.12	Recall against precision plots for datasets CT-NIH, MRI-A and MRI-B using Method 2.	112
6.13	Dice similarity coefficient (DSC) and Jaccard index (JI) box plots for datasets CT-NIH, MRI-A and MRI-B using 3D-Method 2.	113

6.14	Pancreas segmentation results in six different MRI image scans (volumes) using Method 2 and 3D-Method 2. Every column of two 3D pancreas reconstructions highlights the segmentation results and DSC in the same MRI volume using Method 2 (top row) with 3D-Method 2 (bottom row). The segmentation outcome (green) overlaps the ground-truth (red) with computed DSC.	114
6.15	Pancreas segmentation results in four different CT image scans (volumes) using Method 2 and 3D-Method 2. Every column of two 3D pancreas reconstructions highlights the segmentation results and DSC in the same CT volume using Method 2 (top row) with 3D-Method 2 (bottom row). The segmentation outcome (green) overlaps the ground-truth (red) with computed DSC.	114
7.1	Columns (a) (b) and (c) highlights sample slice and corresponding segmentation outcome from MRI-A, MRI-B and CT-NIH respectively. Notice the variation in noise and image distortion between (a) and (b), and the difference in blur and sharpness between (a) and (c). All segmentation results (green) are mapped over the ground-truth (red) and have approximately 85% accuracy in DSC.	124
8.1	Overview of generalised segmentation framework: the initial input is a 3D medical image scan that is processed to segment an organ, muscle or tumour. The segmentation result is processed through a biomedical statistical and classification analysis model, which consequently outputs the final prediction to indicate the likelihood of a condition or progression of a current condition.	131
8.2	Dice similarity coefficient (DSC) and Jaccard index (JI) box plots for liver segmentation in MRI using Method 2.	131
8.3	Dice similarity coefficient (DSC) and Jaccard index (JI) box plots for iliopsoas muscles segmentation in MRI using Method 2.	132
8.4	Dice similarity coefficient (DSC) and Jaccard index (JI) box plots for liver and iliopsoas muscles segmentation in MRI using 3D-Method 2.	132
8.5	Segmentation results in six different MRI image scans (volumes) using Method 2. Every column corresponds to a single MRI volume. From left, first two rows display sample MRI axial slices with segmentation outcome (green) against ground-truth (red), and computed DSC; third row displays 3D reconstruction of entire liver (green) segmentation against its ground-truth (red) with computed DSC.	133

8.6	Liver organ segmentation results in six different MRI image scans (volumes) using Method 2 and 3D-Method 2. Every row of two 3D liver reconstructions highlights the segmentation results and DSC in the same MRI volume using Method 2 and 3D-Method 2 respectively. The segmentation outcome (green) overlaps the ground-truth (red) with computed DSC. .	133
8.7	Segmentation results in six different MRI image scans (volumes) using Method 2. Every column corresponds to a single MRI volume. From left, first two rows display sample MRI axial slices with segmentation outcome (green) against ground-truth (red) and computed DSC; third row displays 3D reconstruction of the entire iliopsoas muscles (green) segmentation against its ground-truth (red) with DSC.	135
8.8	Iliopsoas muscles segmentation results in four different MRI image scans (volumes) using Method 2 and 3D-Method 2. Every column of two 3D iliopsoas muscles reconstructions highlights the segmentation results and DSC in the same MRI volume using Method 2 (top row) and 3D-Method 2 (bottom row). The segmentation outcome (green) overlaps the ground-truth (red).	135
8.9	Dice Score coefficient (DSC) and Jaccard index (JI) box plots for “abnormal” and “normal” kidneys segmentation in MRI using 3D-Method 2.	137
8.10	Dice Score coefficient (DSC) and Jaccard index (JI) box plots for “abnormal” and “normal” kidneys segmentation combined in MRI using 3D-Method 2.	137
8.11	Using 3D-Method 2, a first sample batch of segmentation results in three different DCE-MRI scans (4D volumes) depicting clinically “normal” kidneys. Every column corresponds to a single DCE-MRI 4D volume. Top row displays sample slices with segmentation outcome (green) that overlap the ground-truth (red) and DSC; bottom row displays 3D reconstruction of whole kidneys segmentation that overlaps the ground-truth and DSC.	138
8.12	Using 3D-Method 2, a second sample batch of segmentation results in three different DCE-MRI scans (4D volumes) depicting clinically “normal” kidneys. Every column corresponds to a single DCE-MRI 4D volume. Top row displays sample slices with segmentation outcome (green) that overlap the ground-truth (red) and DSC; bottom row displays 3D reconstruction of whole kidneys segmentation that overlaps the ground-truth and DSC.	139

8.13	Using 3D-Method 2, a first sample batch of segmentation results in three different DCE-MRI scans (4D volumes) depicting clinically “abnormal” kidneys. Every column corresponds to a single DCE-MRI 4D volume. Top row displays sample slices with segmentation outcome (green) that overlap the ground-truth (red) and DSC; bottom row displays 3D reconstruction of whole kidneys segmentation that overlaps the ground-truth and DSC.	140
8.14	Using 3D-Method 2, a second sample batch of segmentation results in three different DCE-MRI scans (4D volumes) depicting clinically “abnormal” kidneys. Every column corresponds to a single DCE-MRI 4D volume. Top row displays sample slices with segmentation outcome (green) that overlap the ground-truth (red) and DSC; bottom row displays 3D reconstruction of whole kidneys segmentation that overlaps the ground-truth and DSC.	140

Glossary

- **Axial:** An axial view refers to a two-dimensional (2D) image taken parallel to the horizontal ground, from top to bottom (xy -plane).
- **Computer-aided diagnosis:** A computer-aided diagnosis (CADx) system processes digitised medical images of radiological scans, in order to highlight areas of organ impairment and disease. Consequently, CADx can provide additional guidance for medical image analysis.
- **Computer tomography:** Computer tomography (CT) is a medical imaging technique that employs a sequence of X-ray images taken from different angles to generate cross-sectional images of the human anatomy. A pre-requisite iodine-based contrast is required, which is injected into the vein of the subject or patient 30 minutes to 1 hour before the scanning procedure.
- **Coronal:** A coronal view refers to a two-dimensional (2D) image taken perpendicular to the horizontal ground, from back to front (xz -plane).
- **Image artefact:** An image artefact is a feature that appears in a radiological scan which is not present inside the original scanned or imaged part of the human body. Artefacts, such as noise, streaks, distortion and blurring, remain problematic in Magnetic Resonance Imaging (MRI) due to magnetic field disturbances, patient motion and signal processing.
- **Image segmentation:** Image segmentation is the process of partitioning a digital image into multiple segments based on a given set of criteria. For example, a target object (e.g., the pancreas) is segmented in a 2D medical image. Each pixel in the image relating to the pancreas is labelled as ‘foreground’ or 1 and each pixel relating to non-pancreas is labelled as ‘background’ or 0.
- **Label:** A label refers to the representation assigned to a single or group of pixels or voxels in a digital image, e.g. a “pancreas” label versus a “non-pancreas” label.
- **Magnetic resonance imaging:** Magnetic resonance imaging (MRI) is a medical imaging technique used to generate radiological images of the human anatomy.

MRI does not involve exposure to ionising radiation. An MRI scanner is a device that employs strong magnetic fields and magnetic field gradient to generate radiological images of the human anatomy.

- **Modality:** Modality describes the type of medical imaging technique that utilises a scanning device to produce images of internal physiological and anatomical structures of a patient. Examples of medical imaging modalities include MRI, CT and ultrasound.
- **MRI sequence:** An MRI sequence corresponds to a scanner imaging protocol, where a particular setting of pulse sequences and field gradients are chosen to emphasise a particular appearance in the resultant image.
- **Pixel:** A pixel or “picture element” refers to the smallest value or referable element in a two-dimensional (2D) digital image.
- **Sagittal:** A sagittal view refers to a two-dimensional (2D) image taken perpendicular to the horizontal ground, from left side to right side (yz -plane).
- **Scanner imaging protocols:** Scanner imaging protocols describe the way in which medical images are acquired, including the choice of scanner modality (e.g., MRI, CT, ultrasound), in addition to how the images are processed for evaluation. Protocols are also dependent on the scanner’s hardware and software, time constraints, patient factors and the radiologist’s preference.
- **T1-weighted:** A T1-weighted image corresponds to an MRI sequence where scanned areas with high fat content appear bright and areas filled with water appear dark.
- **T2-weighted:** A T2-weighted image corresponds to an MRI sequence where scanned areas filled with water appear bright and tissues with high fat content (e.g. white matter) appear dark.
- **Tissue:** A tissue can be described as the grouping of human cells that lie in close proximity and form organs, skin, bone and other parts of the human body.
- **Voxel:** A voxel represents the smallest value or referable element in a three-dimensional (3D) digital image.

Chapter 1

Introduction

1.1 Introduction

The accurate, computer-aided quantitative segmentation and classification of organs can produce valuable information towards analysing organ-related disorders¹, and provide additional guidance towards stratifying subjects after diagnosis or before clinical trials. Research studies highlight that automatic segmentation and computer-aided investigation of significant organ volume variations in patients with conditions² such as polycystic liver disease (PLD)³ [17], renal (kidney) disease [18], and type 1 and 2 diabetes mellitus [12, 13], has played a vital role in raising the quality of biomedical research [19]. For example, an enlarged liver is correlated with PLD subjects, and a larger liver volume is more likely to cause a higher symptom burden [42]. Furthermore, increased kidney volume correlates with early Autosomal dominant polycystic kidney disease (ADPKD) [43]. Last, and certainly more appropriate to the main context of this thesis, the pancreas volume is reduced in patients at the onset of type 1 diabetes [26], and the mean pancreas volume is reportedly 33% less in subjects with type 2 diabetes than in patients who have clinically “normal” glucose tolerance [44]. On a different note, the pancreas volume is restored to near clinically “normal” once type 2 diabetes is reversed [45].

Computer-aided diagnosis (CADx) systems can, therefore, support radiological interpretations of medical scans by providing a “second opinion”, help to reduce possible misdiagnosis, and consequently guide better therapy planning [20]. Recent research literature reports segmentation accuracy scores of 90% or above for the automatic segmentation of organs such as the liver [21], kidneys [22] and spleen [23]. The pancreas,

¹A disorder is a disruption to regular organ structure and function.

²A condition is a state of health that disrupts feelings of regular wellbeing. A disease or disorder can be medical condition.

³A disease describes a condition that impairs regular organ structure and function, characterised by specific signs and symptoms.

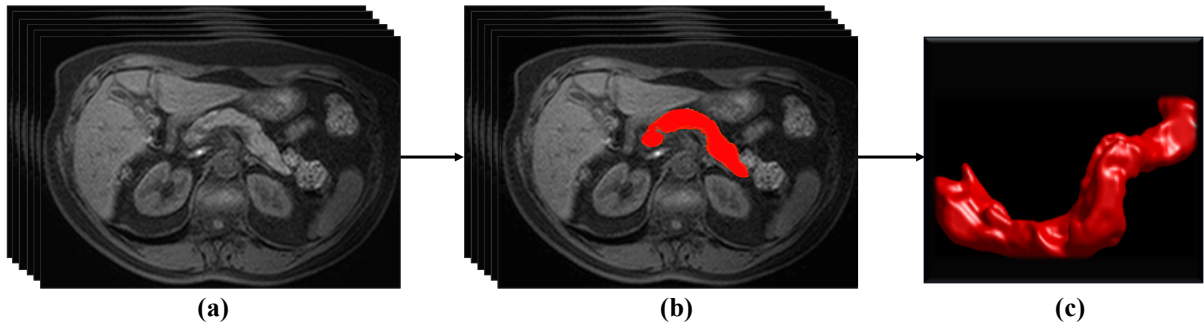


Figure 1.1: Overview of sequence of steps towards automatic pancreas segmentation in Magnetic resonance imaging (MRI) volumes, and 3D rotational reconstruction of the segmented organ: (a) Sequence of original 2D axial image ‘slices’ from an MRI volume; (b) Sequence of segmented (red) 2D image ‘slices’ from the MRI volume; (c) 360-degree 3D digital reconstruction visualisation of the segmented pancreas.

however, presents more significant challenges due to size, structure and location within the abdomen. The pancreas has high structural variability and a full inspection from a scan is problematic since this particular organ lies just behind the stomach: the head of the pancreas touches the small intestine and often overlaps with surrounding abdominal fat, artery and veins. In general, the greyscale (image) intensity of the pancreatic region in a radiological scan, as shown in Figure 1.1(a), is very similar to nearby tissue and consequently increases the challenge of accurate segmentation. Still, despite the presence of different medical image segmentation approaches, a significant number of state-of-the-art methods are restricted to a single modality⁴ or scanner protocol⁵. Furthermore, several methods in research literature produce a relatively high standard deviation for a given number of segmented image volumes, reflecting poor stability [21, 23, 25, 27, 28, 32, 33, 36, 70].

Taking on board the above challenges, the remainder of this chapter introduces the main problem to solve in this thesis and presents the original contributions to knowledge in computer vision that addresses and solves this problem, delivering an extensive application to biomedical research that can be translated to a clinical setting. Section 1.2 delivers the main objective of the research including key highlights of the methods proposed in the PhD project described in this thesis. Section 1.3 introduces the original contributions to computer science (computer vision) in medical image analysis, highlighting the impact of these contributions as elaborated in Chapter 3. Section 1.4 details a list of conference and journal publications, both achieved and under review, within

⁴Modality describes the type of medical imaging technique that utilises a scanning device to produce images of internal physiological and anatomical structures of a patient. Examples of medical imaging modalities include MRI, CT and ultrasound.

⁵Scanner imaging protocols describe the way in which medical images are acquired, including the choice of scanner modality (e.g., MRI, CT, ultrasound), in addition to how the images are processed for evaluation. Protocols are also dependent on the scanner’s hardware and software, time constraints, patient factors and the radiologist’s preference.

the lifecycle of this PhD project. Section 1.5 provides a brief description of remaining chapters. Section 1.6 summarises the objective of research and published methods that are going to be explored, examined and discussed in the following chapters.

1.2 Objective of Research

The technical objective of the PhD project is the development of a computing technique for automatic, quantitative pancreas segmentation in 3D radiological scans (image volumes) that were obtained using different scanner protocols and exhibit high levels of noise diversity. The upcoming Chapters 4 and 5 present two main, principal methods and one method that is an extension of a principal method. The first method is known as Method 1; the second method is known as Method 2, and the extended method is known as 3D-Method 2. The proposed methodologies of Method 1, 2 and 3D-Method 2 are optimised for generalisability and can therefore be applied to other organs, muscular tissue and tumour⁶ segmentation tasks. Moreover, the methodological novelties produce very detailed organ boundary tracing or contouring, which is an essential determinant of morphological features, including organ surface characterisation.

While several reported state-of-the-art methods are limited to a single modality or scanner protocol, and demonstrate poor statistical stability, the proposed methodologies of Method 1 and 2 produce quantitative segmentation accuracies that yield a relatively low standard deviation and reflect robustness to diverse image quality. Accurate and robust automatic quantitative pancreas segmentation, as highlighted in Figure 1.1(b), can produce morphological measures such as volume and curvature⁷, which can support biomedical research establish the severity and progression of a condition, e.g. type 2 diabetes; it can also provide guidance towards stratification of subjects after diagnosis or prior to clinical trials; and help shed additional light on detecting early signs of pancreatic cancer. The impact of accurate quantitative pancreas segmentation, which will be elaborated in Chapter 2, can support radiological interpretations of magnetic resonance imaging (MRI) scans through a 360-degree, 3D digital reconstructed model of the segmented pancreas as shown in Figure 1.1(c).

The project detailed in this thesis aims to harness the advantages of pre-existing classical image processing techniques in computer vision, including graph theory, image intensity thresholding and distribution normalisation, and 3D reconstruction rendering. Furthermore, existing machine learning⁸ algorithms are employed, including structured

⁶A tumour refers to swelling or abnormal tissue growth that forms a mass or lump-like shape.

⁷The curvature of the pancreas as a 3D surface describes the local shape of that surface, which could be interpreted with respect to the curvature range for a clinically ‘normal’ pancreas.

⁸A machine learning algorithm utilises “training data”, in which the algorithm is optimised over time given a finite number of iterations. The “trained” algorithm performs predictions on non-training “test data” without being explicitly programmed.

forest learning (edge detection), random forests (image patch prediction) and artificial neural networks (image contrast parameter prediction). Furthermore, deep learning⁹ architectures are utilised to address the problems caused by (1) high levels of artefacts and noise in MRI; (2) inter-patient variability of the pancreas; and (3) vague organ boundaries in MRI. Combined with popular deep learning architectures for medical image segmentation, a novel training loss function tackles organ boundary detection, and useful data augmentation generates information diversity for improved training of the deep learning model. It is the ambition to rigorously optimise the proposed algorithms to (1) limit computational costs relative to other methodologies in the research literature; (2) expand application and generalisability to image volumes of different spatial dimensions, diverse levels of noise and artefacts, and acquisition using different scanner imaging protocols; (3) produce detailed organ contouring as opposed to an approximate tracing; (4) raise the mean segmentation accuracies relative to the state-of-the-art using diverse evaluation methods that examine the (a) true-positive pancreas predictions versus true-negative pancreas; and (b) the surface or contouring deviation between the resultant segmentation and desired outcome.

1.3 Introducing Original Contributions to Knowledge

Taking into account the growing necessity to accurately delineate or segment organs of high inter-variability, such as the pancreas, from but not exclusively 3D MRI radiological scans, this thesis presents a number of contributions in the field of computer vision with an impact to biomedical research, addressing the growing demand of accurate, quantitative pancreas segmentation. It should be noted that Chapter 3, Section 3.4 will further explain the contributions and respective attributes:

- (a) a learned intensity model enhances the contrast in 3D radiological scans (image volumes) to highlight the pancreas from surrounding non-pancreas tissue;
- (b) a hybrid energy-minimisation segmentation approach exploits edge detection to yield detailed, optimal contouring (delineation) of the pancreas;
- (c) a post-processing stage integrates principal geometric descriptors that characterises pancreas tissue and employs radiological expert-knowledge about anatomical structure for refined tissue classification;
- (d) a deep learning novel loss (error) function based on the modified Hausdorff metric and a sinusoidal component improves true-positive pancreas tissue prediction.

⁹Deep learning is a class of machine learning that employs algorithms, known as neural networks, which are inspired by the functioning of the human brain. A deep learning architecture consists of several layers of neural networks.

A range of attributes using (a)-(d) include:

- enhanced automated pancreas boundary preservation and delineation from an image volume;
- improved application to modalities that extend beyond MR imaging;
- robustness to noise-artefacts in medical image volumes;
- computational efficiency in comparison to reported state-of-the-art;
- and the generalisability of the contributions extends to image volumes acquired using MRI scanner protocols that are not optimised for any particular organ, and thus delivers quantitative segmentation results for other abdominal anatomical structures.

1.4 List of Publications

Summarised below is a brief description of all current publications and a submission under-reivew.

1. Villarini, B., Asaturyan, H., Thomas, E.L., Mould, R. and Bell, J.D. (2017). **A Framework for Morphological Feature Extraction of Organs from MR Images for Detection and Classification of Abnormalities**, 30th IEEE International Symposium on Computer-Based Medical Systems (CBMS 2017).
 - This conference paper describes a computational framework for morphological feature extraction in 3D radiological abdominal scans following expert-led manual organ segmentation. A correlative statistical analysis is performed between organ volume and global 3D curvature using 115 manual pancreas segmentations in abdominal MRI scans.
2. Asaturyan, H. and Villarini, B. (2018). **Hierarchical Framework for Automatic Pancreas Segmentation in MRI Using Continuous Max-flow and Min-Cuts Approach**, 15th International Conference on Image Analysis and Recognition (ICIAR 2018).
 - This conference paper details a methodology for automatic pancreas segmentation in MRI. A dataset, containing 130 MRI scans of adult volunteers, is employed for evaluating the efficiency of the proposed approach. An award was achieved for Springer Best Paper.

3. Asaturyan, H., Villarini, B., Gligorievski, A. (2019). **Morphological and multi-level geometrical descriptor analysis in CT and MRI volumes for automatic pancreas segmentation.** International Journal of Computerized Medical Imaging and Graphics, Elsevier. Journal impact factor: 3.750.
 - This journal paper details the methodology for automatic pancreas segmentation in two diverse MRI datasets containing 132 and 216 scans, and a publicly available CT dataset containing 82 scans. This approach is statistically stable, reflected by lower metrics in standard deviation compared to state-of-the-art approaches.
4. Asaturyan, H., Thomas, L. E., Bell, D. J., Villarini, B. (2019). **A Framework for Automatic Morphological Feature Extraction and Analysis of Abdominal Organs in MRI Volumes.** Journal of Medical Systems, Springer. Journal impact factor: 3.058.
 - This journal paper presents a framework for automatic morphological feature extraction in computer-aided 3D organ reconstructions following organ segmentation in 3D radiological scans. Two different Magnetic Resonance Imaging (MRI) datasets are evaluated to segment the pancreas and liver. Both experiments highlight a negative correlation between 3D curvature and volume with a statistical difference ($p < 0.0001$). Such a tool can support the investigation into organ related conditions such as obesity, type 2 diabetes mellitus and liver disease.
5. Asaturyan, H., Thomas, L. E., Fitzpatrick, J., Bell, D. J. and Villarini, B. (2019). **Advancing Pancreas Segmentation in Multi-protocol MRI Volumes Using Hausdorff-Sine Loss Function.** International Workshop on Machine Learning in Medical Imaging (MLMI) in conjunction with International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI) 2019, Springer.
 - This conference paper proposes a dual-stage automatic segmentation method to address the problem of vague organ boundaries in high class-imbalanced data, by integrating a deep learning-based novel loss function to rigorously optimise boundary delineation. This approach is statistically stable and outperforms state-of-the-art methods on MRI.

6. Asaturyan, H., Villarini, B., Sarao, K., Chow S. K., Afacan, O. and Kurugol, S. (*Under review*). **Improving Automatic Renal Segmentation in Clinically Normal and Abnormal Paediatric DCE-MRI via Contrast Maximisation and Convolutional Networks for Computing Markers of Kidney Function**. Journal of Magnetic Resonance Imaging, Elsevier. Journal impact factor: 2.053.

- This paper presents a fully modular framework for automatic renal parenchyma segmentation, including medulla and cortex, in 4D Dynamic Contrast Enhanced (DCE) Magnetic Resonance Imaging (MRI) volumes. The proposed framework is evaluated on a paediatric dataset containing 60 diagnosed 4D DCE-MRI volumes that show varying conditions affecting kidney function; and it achieves quantitative segmentation results that outperform a state-of-the-art approach and demonstrates statistical stability.

1.5 Organisation of Remaining Chapters

This section provides a summary of each remaining chapter throughout this PhD thesis. Chapter 2 provides the background, rationale, motivation and challenges that drives the development of this project. Chapter 3 provides a summary of current state-of-the-art approaches that produce promising segmentation results in medical image scans. Furthermore, this chapter references baseline architectures that are built upon in proceeding methodology-based chapters. Chapter 4 (Method 1) presents the first of two methodologies for automatic pancreas segmentation: this chapter introduces a novel intensity model for digital image contrast enhancement in volumes that aid the segmentation contouring outcome at a later processing stage. Furthermore, this chapter presents a novel refinement stage that integrates both expert radiological knowledge with computer-aided analysis of distinct segmentations for tissue classification, i.e. “pancreas” versus “non-pancreas”. Chapter 5 (Method 2) delivers the second of two principal, proposed methods for automatic pancreas segmentation: this method exploits the advantages of current deep learning architectures and integrates a novel loss function for advancing the accuracy prediction rate in comparison to a readily used conventional loss function. Combined with novelties described in Chapter 4, the second proposed method ensures generalisability to other organ and muscular tissue segmentation tasks. This chapter also investigates and compares the potential advantages and disadvantages of a 3D deep learning architecture (3D-Method 2) to Method 2. Chapter 6 presents, discusses and compares the segmentation evaluation results obtained in Method 1 (Chapter 4) and Method 2 (Chapter 5) using two MRI datasets and one CT dataset. This chapter also details the numerical implementation for Methods 1 and 2, and 3D-Method 2, in ad-

dition to key technical tools and libraries. Chapter 7 critically compares and discusses the quantitative segmentation results obtained using Methods 1 and 2, and 3D-Method 2 with reported state-of-the-art methodologies in terms of quantitative and qualitative accuracy. This chapter also explores potential method optimisations in Method 2 and 3D-Method 2 to boost segmentation performance. Chapter 8 discusses the generalisability of Method 2 highlighting the impact to biomedical research and clinical studies, including significant correlative analysis between computed morphological features and identification of important prognostic biomarkers in 3D radiological scans. This chapter also explores an extended application, in which the input of interest consists of image volumes from multiple imaging protocols (sequences) from the same type of scanner or multiple scanner modalities. Last, but certainly not least, Chapter 9 delivers a summary of key challenges, motivations and achievements accomplished using the proposed methods in this PhD research project. This chapter also reexamines the various drawbacks of the reported state-of-the-art and presents outperforming quantitative results using the methods that are proposed in this thesis, and reveals the significance of employing the original contributions to knowledge into future work.

1.6 Conclusion

There is a significant need for accurate organ segmentation with a growing dependency on computer-aided systems as support-tools for detection and diagnosis from 3D radiological scans. The accurate organ boundary delineation in a scan can provide important clinical information when presented as a 360-degree, 3D digital reconstruction, or when evaluating the scan on a 2D-by-2D (slice-by-slice) basis. The progression of research objectives in the lifecycle of this PhD project is reflected in published conference and journal papers, highlighting the investigation of classical image progressing methods through to deep learning technologies that optimise the contribution to knowledge in the field of medical image analysis, and the application of computer science in biomedical research with a clinical translation.

Chapter 2

Background and Rationale

2.1 Introduction

In recent decades the global prevalence of diabetes among adults has risen from 4.7% in 1980 to 8.8% in 2017 [185]. Diabetes, which is a major cause of blindness, heart attacks and stroke, is on the rise and projected to reach 642 million adults by 2040 [209]. Furthermore, pancreatic cancer is one of the leading causes of cancer mortality and one of the most dangerous tumours across the world. Although certain risk factors have been identified such as type 2 diabetes and obesity, the causes are still insufficiently known despite growing research in biomedicine [187]. For this reason, insight into the relevant organ of interest (i.e. the pancreas) is vital, and one way is to extract or segment the pancreas from 3D radiological scans. With that said, there are a number of challenges and limitations that need to be addressed, tackled and minimised in order to ensure an accurate, clinically acceptable segmentation.

Automatic pancreas segmentation in 3D radiological scans is an essential, yet challenging task. As a prerequisite for computer-aided diagnosis (CADx) systems, accurate pancreas segmentation could generate both quantitative and qualitative information towards establishing the severity of a condition and thus provide additional guidance for therapy planning. This chapter discusses the medical conditions affecting the pancreas, and the challenges that limit as well as drive the development of accurate pancreas segmentation methodologies, particularly in multiple modalities including Computer tomography (CT) and Magnetic resonance imaging (MRI) scans. Section 2.2 discusses the significance of pancreas segmentation with focus on improved detection, diagnosis and stratification of patients with diabetes and pancreatic cancer. Section 2.3 presents key challenges involved in producing large-scale accurate segmented pancreas volumes in radiological scans. Section 2.4 summarises this chapter with key highlights regarding medical pancreas conditions and respective MR imaging, and the challenges to interpretation that will be necessary to consider in proceeding chapters.

2.2 Significance of Pancreas Segmentation in CADx

The computation of pancreas measurements, such as volume and curvature, can provide insight into the progression of type 2 diabetes mellitus [12, 13] and assist in detecting pancreatic neoplasms¹ [14]. It is noted that the curvature characterises the surface of the pancreas in terms of its level of “smoothness” or “raggedness”, providing an indication about the potential deformity of the organ from a clinically “normal” pancreas. Indeed, studies have reported that variations in the pancreas contour can be linked to ductal adenocarcinoma [10], and enhanced contour analysis can facilitate stratification of clinically “normal” variations against pancreatic tumours [15]. Obtaining such information firstly requires segmenting the pancreas in radiological image volumes, such as Computer tomography (CT) and Magnetic resonance imaging (MRI). Although manual segmentation can produce very accurate results, it is time-consuming, sometimes prone to inter-observer variability, but above all, it is challenging to replicate given the growing number of available medical image volumes. On the other hand, CADx systems that generate automatic, accurate pancreas segmentation on a scale of thousands of abdominal medical scans could support clinical studies to establish essential correlations between organ volume, curvature and anthropometric² measures [16].

As the upcoming sub-sections detail the essential task of imaging the pancreas for disease detection and diagnosis using both MRI and CT scanning, the paragraph below provides an introductory description about the differences between CT and MRI in terms of organ visibility and image artefacts.

How does MRI and CT Quality Differ?

Despite CT scanning taking less time and costing less than MRI scanning, some types of cancers or lesions can be harder to identify in a CT scan where otherwise detailed in an MRI scan. MR imaging delivers sharp views of organs within the abdomen and soft tissues, joints, and bones of the body. In fact, MRI scanning is often used when CT has failed to deliver sufficient information for disease detection or diagnosis purposes. With that said, MR imaging may not always differentiate between excessive fluid edema (swelling) and cancerous tissue due to similar greyscale intensity and structure [71]. Also, during the MR imaging acquisition stage, signal processing can cause streak artefacts that resemble multiple visible lines across the resulting image. Furthermore, both MRI and CT scanning can produce blurring motion-based artefacts resulting from breathing, cardiac movement and blood flow. Please refer to Appendix A to view a set of 2D images (slices) taken from MRI and CT volumes.

¹A neoplasm refers to abnormal excessive growth of tissue, commonly known as a tumour.

²Anthropometric measures refers to dimensions of the human body such as height, weight, and the size and shape of bone, muscle, and fat tissue.

Often, the first choice of scanner is heavily dependent on the patient’s health history and symptoms. For example, it is often reported that patients with artificial metal joints and a pacemaker avoid MRI as the scanning procedure employs powerful magnetic fields that could cause problems for the patient [214, 215]. In contrast, since CT scanning requires a small dosage of ionising radiation, it has been reported that pregnant women in need of abdominal imaging may risk exposing their unborn baby to radiation during the scanning procedure [216, 217].

2.2.1 Monitoring Diabetes

Diabetes mellitus (DM) or diabetes is a condition that affects the body’s ability to use sugar (glucose) from carbohydrates in the food that is consumed. There are three major types of diabetes, including type 1 diabetes, type 2 diabetes and gestational diabetes [62]. With type 1 diabetes, the pancreas produces very little or no insulin³ due to the body’s immune system destroying the cells that generate insulin in the pancreas. This type of diabetes is most commonly diagnosed from infancy to the late 30s. With type 2 diabetes, the body rejects the effects of insulin or does not produce sufficient insulin that the body can use to support healthy glucose levels. In this case, the cells become resistant to insulin, triggering the release of more insulin to keep blood glucose levels within a normal range. Eventually, the pancreas can deteriorate from producing extra insulin. Type 2 diabetes is, by far the most common type of diabetes amongst adults. Gestational diabetes is a type of diabetes that is triggered by pregnancy. Pregnancy can, to some point, lead to insulin resistance, and this type of diabetes is often diagnosed in middle or late pregnancy. Since high blood sugar levels in an expectant mother are circulated through the placenta to the baby, gestational diabetes must be controlled to safeguard the growth and development of the baby [63].

Both CT and MRI scanning, amongst other imaging modalities, provide significant information about the progress of diabetes and the correlation between other anatomical and physiological factors. For example, the CT and MRI studies described in [60] demonstrate a correlation between diabetes, cerebral atrophy⁴ and lacunar infarcts⁵ but no consistent relation with white matter lesions⁶. The magnetic resonance spectroscopy

³Insulin is a hormone made by the pancreas that allows the body to use sugar (glucose) from carbohydrates in food that is consumed. Insulin helps to keep the body’s blood sugar level from getting too high or too low.

⁴Cerebral atrophy describes the loss of neurons (cells) in the brain.

⁵Lacuna infarcts, also known as Small vessel disease, are small infarcts (2 to 20 mm in diameter of area of dead tissue) in brain matter, resulting from the blockage of a small artery that burst while providing bloody supply to the subcortical areas of the brain.

⁶A lesion refers to damaged or abnormal change in organ tissue. White matter lesions are areas of damaged nerve cells found in the white matter of the brain.

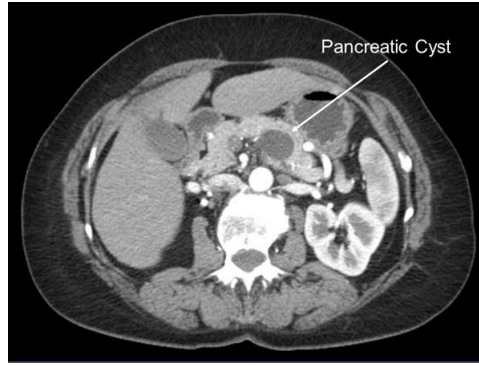


Figure 2.1: A pancreatic cyst, marked by an arrow, as shown on an MRI. Only a small percentage of such cysts lead to cancer [188].

(MRS) studies report raised the ratio of myo-inositol⁷ to creatine⁸ and reduced the ratio of *N*-acetylaspartate⁹ to creatine in patients with diabetes (of either type 1 or type 2). Another CT scanning study demonstrated that diabetic patients had increased lobulation¹⁰ of the pancreas [61]. All parts of the pancreas were inclined to be smaller in patients with diabetes, but the degree of decline was mixed; the decline was ‘modest’ for patients who were not treated with insulin and the decline was more ‘pronounced’ in patients who were dependent on insulin. In general, the size of the pancreas’ body was significantly reduced in all three groups, whereas the size of the pancreatic head was preserved in patients who were not treated with insulin. CT scanning reveals that although the pancreas density in diabetic patients is clinically “normal”, there is increased lobulation. Reduction in size involves the body of the pancreas more than other parts of the gland and is greater in diabetic patients treated with insulin. Therefore, CT scanning of the pancreas might be useful in predicting which diabetic patients will require insulin treatment.

In various other studies, MRI has demonstrated to be a reliable method for measuring pancreatic volume and levels of fat surrounding the pancreas. In a study reported in [64] effective MRI scanning was performed to monitor pancreatic atrophy¹¹ in type 1 diabetes. Four different sequences¹² were employed, including conventional T1-weighted¹³,

⁷Myo-inositol is a carbocyclic sugar that is abundant in the brain. It is a sugar alcohol with half the sweetness of table sugar.

⁸Creatine is an organic acid that occurs naturally and primarily in the liver and kidneys. Its main role is to assist in recycling the energy of cells in mainly muscle and brain tissue.

⁹*N*-acetylaspartate (NAA) is the second-most-concentrated molecule in the brain and is detected in the adult brain in neurons.

¹⁰A lobulation is an appearance resembling lobules. A lobe is a clear anatomical division or extension of an organ.

¹¹Pancreatic atrophy is the permanent deterioration of the pancreas.

¹²An MRI sequence corresponds to a scanner imaging protocol, where a particular setting of pulse sequences and field gradients are chosen to emphasise a particular appearance in the resultant image.

¹³T1-weighted image: tissues with high fat content appear bright and areas filled with water appear dark.

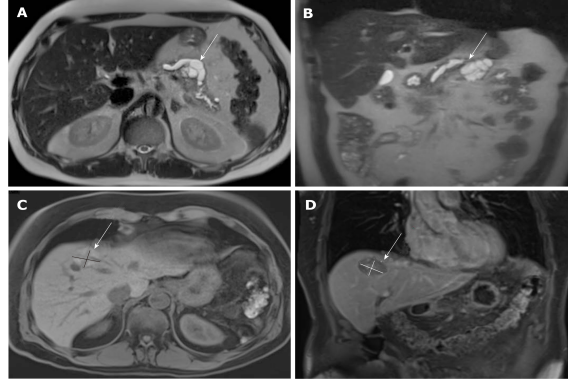


Figure 2.2: MRI scans of tumour lesions. A (T2 axial view) and B (T2 coronal view) highlight a primary tumour lesion 2 months before the initial diagnosis, which identifies a cystic lesion at the location between body and tail of the pancreas with a size of 43 mm \times 38 mm \times 30 mm (A) and dilatation of the pancreatic duct of 13 mm (B). C (T1 axial) and D (T1 coronal) highlight a relapsed hepatic lesion 20 months after an after initial diagnosis [189].

conventional T2-weighted¹⁴, volumetric interpolated breath-hold examination (VIBE)¹⁵ and T1-weighted breath-hold with fat suppression (T1BHFS). From here, the pancreas organ was recognised on coded images by an observer and volumes were estimated by interpolation. Consequently, these experiments showed that MRI could be used to determine the natural history of pancreatic atrophy in diabetes. In another study, fat levels in patients with type 2 diabetes (pancreatic triglyceride) were measured using a targeted MRI scanning method. This pancreas-specific method led to the following conclusion: for a person with type 2 diabetes that has a pancreas volume of 50 ml, this is the equivalent to around 0.6 grams of fat. On a different note, the patients who had never had diabetes saw no change in the level of fat in their pancreas, demonstrating that increases in pancreatic fat are specific to people who develop type 2 diabetes [65]. On the topic of pancreas volume measurements, another study compared pancreas volume (PV) measurement using MRI-based planimetry in patients with type 2 diabetes, to PV in normoglycemic¹⁶ individuals [12]. The pancreas contours of 32 type 2 diabetics and 50 normoglycemic individuals were outlined on non-gadolinium¹⁷ (T1-weighted) 3D fat suppressed gradient-echo images. Afterwards, the PV index (PVI) was calculated as PV/weight to adjust PV for each patient’s weight. PVs and PVIs in both groups were compared using t-tests and regression models corrected for weight, age and sex. The

¹⁴T2-weighted image: scanned areas filled with water appear bright and tissues with high fat content (e.g. white matter) appear dark. This is good for demonstrating pathology since many lesions are associated with an increase in water content.

¹⁵Volumetric interpolated breath-hold examination (VIBE) sequence produces T1-weighted images using interpolation. VIBE is a modified form of fast low angle shot Magnetic resonance imaging (FLASH) sequence allowing high-resolution imaging in less than 30 seconds breath-hold.

¹⁶Normoglycemic implies having a clinically “normal” amount of glucose in the blood.

¹⁷Gadolinium-based contrast agents, given by intravenous injection, are a widely used for MRI scanning.

study concluded that PV is reduced in type 2 diabetes compared to normoglycemic individuals and can be measured using MRI without contrast injection. Expanding on the ability of MRI to provide detailed information about a pancreatic condition, a study [66] has reviewed magnetic resonance cholangiopancreatography (MRCP) findings in patients with diabetes, pancreatic exocrine insufficiency, and with combined pancreatic exocrine insufficiency and diabetes. This particular report showed that chronic pancreatitis MRCP findings were present with increasing frequency in groups of diabetes and insufficient pancreatic exocrine.

2.2.2 Pancreatic Cancer and Pancreatitis

Pancreatic cancer is caused by the abnormal and uncontrolled growth of cells in the pancreas. In fact, during the early stages, a pancreatic tumour or neoplasm does not usually cause any symptoms, which can make it difficult to diagnose. Pancreatic neoplasms can originate in the exocrine-based¹⁸ or endocrine-based¹⁹ functioning period; such tumours that stem in the pancreas can be malignant epithelial²⁰ or otherwise, and can resemble cystic lesions. One such neoplasm, ductal pancreatic adenocarcinoma, constitutes roughly 90% of all malignant neoplasms [210, 211] in which the tumour is located within the pancreas in up to 70% of cases. Approximately 80% of the tumours occur in patients aged 60 to 80 years old, in which men are twice as likely affected than women [212]. Furthermore, being able to differentiate between primary cystic neoplasms from pancreatic ductal adenocarcinomas and other pancreatic malignancies is essential for therapy planning. Interestingly, a study described in [51] reported that CT was insensitive towards differentiating cystic tumours whereas MRI displays a well-defined boundary of such tumours.

Early detection of pancreatic cancer is paramount to improving patient survival. More often than not, patients who undergo abdominal scanning are diagnosed as having a pancreatic cyst²¹ (see Figure 2.1), and yet not all pancreatic cysts will progress to pancreatic cancer. For this reason, effective computational tools to detect or differentiate a pancreatic cancer arising within a pancreatic cyst early would be a valuable asset to the clinical community. For example, Figure 2.2 shows a follow-up case study MRI scan about one year after initial symptoms were detected in a CT scan, in which multiple lesions were seen in the tail of the pancreas (Figure 2.2 A and B). A splenopancreatectomy was performed (removing the spleen and the pancreas) and a tumour of 7 cm × 5 cm × 3.5 cm on the pancreatic tail was surgically removed.

¹⁸Functioning as an exocrine gland, the pancreas excretes enzymes to break down the proteins, fats and carbohydrates in food.

¹⁹Functioning as an endocrine gland, the pancreas secretes the hormones insulin and glucagon to control blood sugar levels.

²⁰Malignant epithelial means in relation to thin tissue forming the outer layer of an organ surface.

²¹A cyst is a sac of tissue filled with air, fluid or other substances.

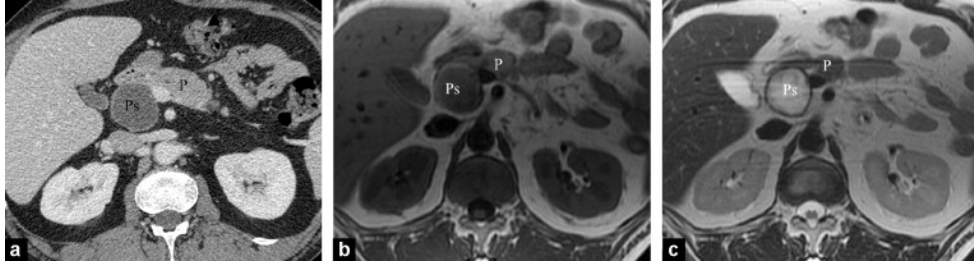


Figure 2.3: Pancreatic pseudocyst (Ps) in a man in his early 40s known with gallstones [190, 193]. (a) CT axial (b) T1-weighted (c) and T2-weighted MRI, obtained 6 weeks after the beginning of an acute pancreatitis condition, highlights a homogenous-kind fluid collection (Ps) with a fibrous capsule pancreas (P).

Acute Pancreatitis

Acute pancreatitis is an acute inflammatory process of the pancreas; in cases where acute pancreatitis is severe, imaging is performed to examine the presence of peripancreatic²² fluid collections and assess the pumping of blood in pancreatic parenchyma²³. The standard imaging modality of choice in acute pancreatitis is CT due to its accuracy and availability [53, 54]. That said, it has been reported that MRI scanning can also fulfill these objectives and have recently been suggested as an substitute to CT for the initial examination of acute pancreatitis [40]. For example, Figure 2.3 compares a CT, T1-weighted and T2-weighted MRI scan on the same abdominal region in the same patient. In the evaluation of acute pancreatitis, MRI can accurately show the presence and extent of necrosis²⁴ and peripancreatic fluid collections, and is a better alternative to CT in detecting mild acute pancreatitis [55].

A routine pancreas protocol including MRI fat-suppressed T1-weighted sequences and a series of T1-weighted gradient recalled echo (GRE) sequences prior to and post-gadolinium is a suitable arrangement for examining acute pancreatitis and reaching an informed opinion [54]. In the case of severe acute pancreatitis, gadolinium-enhanced MRI is helpful for evaluating pancreatic parenchymal blood circulation. Moreover, magnetic resonance pancreatography (MRP), which is a technique that focuses on evaluating of the pancreatic duct²⁵ alone, or magnetic resonance angiogram (MRA), opens the way for accurate diagnosis of complications affecting surrounding blood vessels.

²²Peripancreatic means tissue occurring in or surrounding the pancreas.

²³Pancreatic parenchyma is the functional part of the pancreas in the body.

²⁴Necrosis is the death of majority (or all) the cells in an organ (or tissue) due to disease or failure of the blood supply.

²⁵A duct refers to a constrained passage or channel leading from an exocrine gland or organ.

Chronic Pancreatitis

Chronic pancreatitis can be described as fibrosis²⁶ and cellular infiltration²⁷ [56] of the pancreas. Chronic pancreatitis is an irreversible disorder where pancreatic exocrine and endocrine cease to function. An MRI examination that screens for chronic pancreatitis will generally include MRP, as well as T1-weighted and T2-weighted images. The signal intensity of T1 and T2-weighted images tends to decrease with the presence of what is considered chronic pancreatitis.

Secretin-administered²⁸ MRCP imaging presents the ductal anatomy, highlighting the severity of pancreatic duct obstruction, and possible exocrine damage, all of which may provide useful guidance prior to treatment planning. On a different note, a downside of MRI scanning may include poor sensitivity towards detecting scattered calcifications²⁹ within the pancreas or in the distal³⁰ portion of the pancreatic duct, and thus MRI scanning might not help to detect the cause of non-calcified³¹ defects.

The ability to detect mild chronic pancreatitis and differentiate from pancreatic adenocarcinoma remains challenging for non-invasive imaging methods. At present, endoscopic retrograde cholangiopancreatography (ERCP) is the most sensitive imaging procedure for detecting the influence of pancreatic duct side branch in the early stages of chronic pancreatitis [58]. CT imaging is the first choice for screening for advanced or severe chronic pancreatitis. In some cases, non-enhanced CT is combined with MRCP in order to examine disease before treatment planning effectively. Due to the non-invasiveness of the imaging technology, MRI is suitable for follow-up screenings in patients with chronic pancreatitis after cure from ductal obstruction is obtained with endoscopic treatment [57].

2.3 Challenges

This section explores the various challenges surrounding accurate pancreas segmentation in medical image scans, not least the interobserver variability in pancreas contouring (boundary tracing) and tumour volume measurement. The differences in imaging modalities (e.g. CT vs MRI), image quality and limitations of shared data are also addressed.

²⁶Fibrosis is the formation of excess fibrous connective tissue in an organ in a reactive process.

²⁷Cellular infiltration is the movement of cells from their origin (or the direct extension of cells) into organs or tissues as a consequence of unusual cell growth.

²⁸Secretin is a hormone produced by the digestive tract, and often used as a medicine.

²⁹Calcification is the accumulation of calcium salts in a body tissue.

³⁰Distal refers to part of an anatomy situated farthest from point of attachment or origin.

³¹Non-calcified nodules (swelling or aggregation of cells in body) are classified as ground-glass opacities (image findings that indicate a partial filling of air spaces)

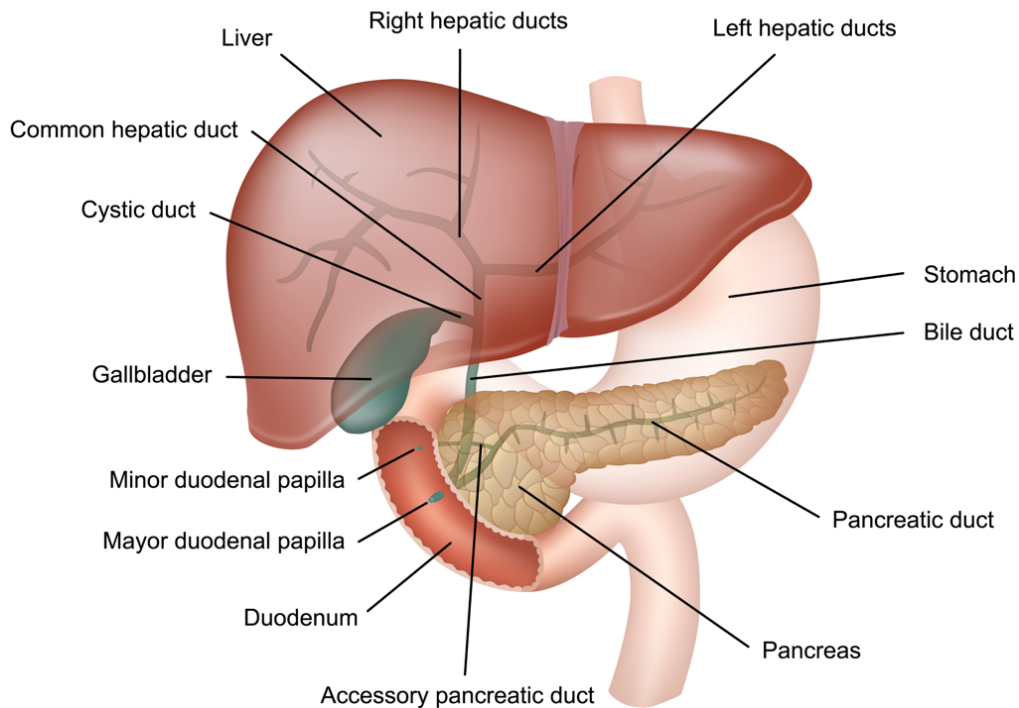


Figure 2.4: Location of pancreas in abdominal cavity [158]. The pancreas lies on the posterior abdominal wall. Notice that the pancreas is located just behind the stomach in the back of the upper left abdomen. The head of the pancreas lies within the duodenum curve, which is the first section of the small intestine. The tip of the pancreas extends across the abdominal cavity.

2.3.1 Structural Variability

Pancreas segmentation, compared to other abdominal organs, presents more significant challenges due to size, structure and location. Firstly, the pancreas has high structural and inter-patient variability [94], and the location within the abdominal cavity shifts from patient to patient. Secondly, a full inspection from a scan is problematic since the organ lies just behind the stomach, as shown in Figure 2.4. The head of the pancreas touches the first part of the small intestine (duodenum) and often overlaps with surrounding abdominal fat and vessels. The high structural variability of the pancreas and respective abnormal cell growth has led to considerable interobserver variation in delineation of tumours in CT volumes amongst expert radiation oncologists [95]. The mean overall observer variation (root-mean-square) was 0.63 cm for pancreatic gross tumour volume (GTV) and 0.80 cm for internal GTV (accounting for the respiratory motion). Another study has highlighted over 40 variations in the lateral contour of the clinically “normal” head and neck of the pancreas in CT image volumes for 119 patients [96]. For example, approximately 35% of patients had discrete lobulations of pancreatic tissue greater than 1 cm lateral to the superior pancreaticoduodenal artery³².

³²The superior pancreaticoduodenal artery is an artery that supplies blood to the duodenum and pancreas.

2.3.2 MRI vs CT Modality

For an initial pancreas examination, a patient normally undergoes CT scanning [39]. If results are inconclusive and further information is required, they are often referred to an MRI examination or an ERCP evaluation. Unlike a CT procedure that requires an ionising contrast media, a commonly known advantage of MRI includes the non-ionising radiation feature³³. Conventional contrast-enhanced CT scanning does, however, produce images of high spatial resolution compared to MRI. Furthermore, a CT is more favourable compared to MRI for patients who are extremely ill. The strong magnetic field of an MRI scanner means that patients must notify their physicians of any form of medical implant before a scan. Also, an MRI scanner creates a low, drumming noise, more commonly referred to as “clicking and beeping” and thus patients often require special ear protection. Furthermore, nerve stimulation is created, which is a twitching sensation resulting from the quickly switched magnetic fields in the MRI machine [40].

In some cases, non-contrast CT can be used to evaluate the pancreas in patients with renal failure or patients who exhibit allergic reactions to the (iodinated) conventional contrast agent. As the pancreatic tumours or lesions can only be visualised with contrast imaging, non-contrast CT scans exhibit poor quality for pancreatic tumours and therefore cannot be relied upon to make an accurate diagnosis [38]. Similarly, although conventional MRI does not require contrast agents, there are cases where administration of gadolinium is required. When a tumour is not identifiable on a CT scan, an MRI examination of the pancreas is performed with intravenous administration of contrast, with gadolinium being the most commonly used agent. Identification of pancreatic tumours is hypersensitive³⁴ on gadolinium-enhanced (T1-weighted) because it is lacking sufficient blood vessels.

Imaging Quality

Studies reported in [93] shows that statistical differences were found between pancreatic gross tumour volume (GTV) from different imaging modalities, including MRI fusion and CT. The study reports that the volumes of “at-risk” pancreata based on MRI are generally smaller than those based on CT, and therefore further studies are required to identify the optimal imaging modality or sequence to define GTV. Differing from CT imaging, the low resolution and slower imaging speed of MRI presents additional edge-based artefacts [24]. Figure 2.5(a) displays a single axial slice from an abdominal CT scan. The top image shows the original slice, and the bottom image highlights the ground-truth of the pancreas contoured in red. Figures 2.5(b)(c) display axial MRI slices

³³Non-ionising radiation is a type of electromagnetic radiation that does not carry enough energy to damage human tissue.

³⁴A hyperintensity is an area in an MRI scan that appears lighter in colour than the surrounding tissues, and thus a hypointensity would be darker in colour.

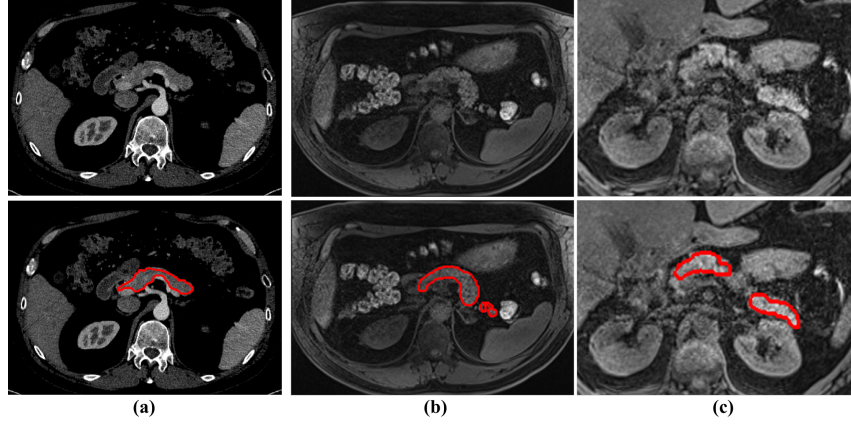


Figure 2.5: The expert-led manual (ground-truth) segmentation of the pancreas is contoured in red. Column (a) displays an axial slice in a CT scan. Columns (b) and (c) display two axial slices from two different MRI scans. Notice the finer image quality of CT in comparison to MRI that is significantly coarse and suffers from greater blurred boundaries between the pancreas, and duodenum.

from two different scans for two different subjects. Notice the similarity in the greyscale intensity of the surrounding tissue. Notice that in column (c), the slices appear “zoomed-in” to emphasise the variation in the visibility of pancreatic features in a slice-by-slice inspection.

2.3.3 Dataset Limitations

In recent years, medical data - and in particular imaging data - has become available for academic research purposes [87, 88, 89, 90, 91, 92]. A number of different organisations provide a platform where researchers can obtain image data to advance their work in medical image computing and validate the effectiveness of their algorithms. That said, there are still many limitations including imaging modality restriction, dataset size and the scanned region of anatomy. For example, the Cancer Imaging Archive (TCIA) provides a publicly available resource of abdominal CT scans [81] but the same cannot be said about MRI modality.

Despite a growing number of “Biobanks” currently provide a platform [85, 86] of medical imaging data for research purposes, a shared platform is yet to be available where researchers can access the same data (especially MRI) and make direct comparison with other methodologies. Also, with increasing scrutiny about “data protection and distribution”, Biobanks that choose to resell data to third parties can generate tensions very similar to those that social networking platforms have faced in recent years [84]. The authors of many publications, with the exception of a publicly available CT data provided by TCIA, rely on data provided by their internal departments or hospitals that collaborate with their department. Also, the data utilised for research purposes

lies in multiples of 10 as opposed to multiples of 100s (or 1000s) as in the case of publicly available natural images databases ImageNet [82] or Pascal VOC [83].

One approach to addressing dataset limitations – and a very popular approach for medical image segmentation tasks – is to create artificial data or “data augmentation” by performing geometrical transformations or warps on the limited data that is available for developing an algorithm. The significance of data augmentation is elaborated in Chapter 5 (Section 5.2.2).

2.4 Conclusion

The severity of pancreatic conditions has propelled researchers to evaluate a medical image scan using different modalities and different scanner imaging protocol sequences, in view to minimise the human inter-observer error and maximise a correct detection or diagnosis. Such motivations are limited by imaging quality, and also, by the high variation of the pancreas’ anatomical structure and positioning. For this reason, there is growing interest in the advancement of computational tools that segment the pancreas from 3D radiological scans and compute morphological features. However, the development of these computational algorithms for medical image analysis is hindered by limitations to open-source datasets in order to compare the effectiveness of different algorithms. It is therefore essential to explore, investigate and develop effective computational methods that exploit the pancreatic anatomical information present in available medical image scans.

Chapter 3

Related Work and Contributions

3.1 Introduction

This chapter provides an overview of state-of-the-art methodologies that drive pancreas segmentation tasks, and moreover, the novel contributions delivered in this PhD project. Section 3.2 highlights current automatic methods in the research literature for pancreas or multiple-organ segmentation, including methods that utilise multi-atlas approaches and convolutional neural networks (CNNs). This section also introduces three baseline CNN architectures that serve as a guide for the methods proposed in Chapters 4 and 5. Section 3.3 builds upon the previous section, discussing a selection of diverse, state-of-the-art methods that predominantly involve CNNs for medical image pancreas segmentation. Section 3.4 provides an overview of the contributions to computer science (computer vision) in medical image analysis as detailed in Chapters 4 and 5, in which the attributes and impact of the contribution applies to other abdominal organs and muscular tissue and can affect the usage of computational methods in a clinical setting.

3.1.1 Quantitative Measure of Accuracy

The resultant segmentation accuracies described in this section are based on two popular statistical metrics that are readily employed in medical image segmentation tasks: Sørensen–Dice coefficient or Dice similarity coefficient (DSC) and the Jaccard index (JI). Both metrics can be described as the quotient of similarity between 0 and 1 (or 0 to 100%), and the formula for each metric is defined in Chapter, 6 Section 6.4. Moreover, it is noted that a *true-positive* and *false-positive* label refers to the correct and incorrect predicted segmentation label representing the pancreas in an image volume, respectively. Similarly, a *true-negative* and *false-negative* label refers to the correct and incorrect predicted segmentation label representing the background (i.e. non-pancreas) in an image volume, respectively.

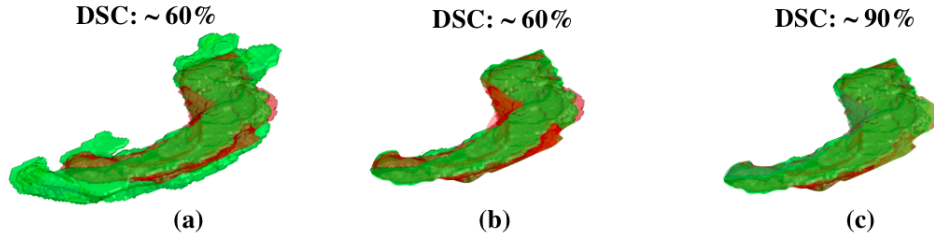


Figure 3.1: Three different segmentation outcomes (green) overlapping the same ground-truth (red) representing a 3D pancreas reconstructed model. Column (a) suffers from excess false-positive segmentation labels; column (b) suffers from false-negative labels; column (c) has significantly less *both* false-positive and false-negative labels segmentation labels than (a) and (b).

It is noted that the JI measures the overlap between the segmentation outcome and desired outcome (i.e. the ground-truth or expert-led manual annotation of the organ), often referred as the size of the intersection between two sets (i.e. the segmentation and ground-truth) divided by the size of the union between these two sets. The DSC, which is similar to JI, describes twice the number of elements common to both sets (i.e. segmentation and ground-truth) divided by the sum of the number of elements in each of these sets. This particular metric function retains sensitivity in more dissimilar data and assigns less weight to outliers [175]. The DSC is different from the JI which only counts true-positives once in both the numerator and denominator. Essentially, unlike the JI, the DSC function does not possess the property of triangle inequality¹.

To illustrate how segmentation accuracies can differ, Figure 3.1 displays three different segmentation outcomes (green) overlapping the same ground-truth (red) representing a 3D pancreas reconstructed model. The segmentations in Figures 3.1 (a) and (b) achieve $\sim 60\%$ in DSC with the former suffering heavily from false-positive and the latter suffers from false-negative labels. Figure 3.1 (c), in contrast, delivers a more accurate DSC score of $\sim 90\%$ while minimising *both* false-positive and false-negative labels.

3.2 State-of-the-art Methods

In recent research literature, various organ segmentation methods have been proposed in the scheme of multi-atlas label propagation approaches [21, 23, 25, 27, 28, 32, 33, 36, 70] and convolutional neural networks [1, 14, 24, 30, 130, 133, 134, 136, 141, 171, 178, 180]. The common factor amongst all of these methods is a reliance upon expert-led manually annotated organs in medical image volumes. Also, most segmentation methods have been performed on CT modality as opposed to MRI, which presents additional difficulties of image artefacts and greater blurred boundaries between organs.

¹Triangle inequality: for any triangle, the sum of the lengths of any two sides must be greater than or equal to the length of the remaining side.

3.2.1 Multi-atlas Label Propagation and Statistical Shape Modelling

Multi-atlas label propagation (MALP) segmentation methods employ “atlases”, which are labelled regions of interest (e.g. organs) and corresponding intensities in multiple radiological image volumes. Such methods usually employ image registration² to align the atlases to an (unseen) test image volume [34]. Label fusion strategies are then used to combine labels from multiple atlases to determine the overall segmentation [35]. Another set of approaches that are often, but not always, coupled with MALP is statistical shape modelling [36]. Such methods define a template shape for a given structure representing the organ of interest, with reliance upon control points along the boundary, after which the entire shape is deformed to match the test image volume.

For example, the approach reported in [33] performs multi-organ segmentation by combining spatial interrelations with probabilistic atlases and incorporating prior knowledge into the model using shape representations of multiple organs. This approach achieves a mean Dice similarity coefficient (DSC) of 46.6% with a standard deviation of 8.7% in a dataset containing 28 CT volumes.

The approach described in [32] firstly registers triple-phase³ CT image volumes together for a particular subject and then performs registration to a reference subject by a landmark-based deformable model. The abdominal cavity is isolated by minimising the liver, spleen, splenic and superior mesenteric veins that lie very close to the pancreas, as shown in Figure 3.2. From here, segmentation is performed via a generated patient-specific probabilistic atlas-guided approach followed by intensity-based classification and post-processing. The proposed method yields a Jaccard index (JI) of 57.9% using 98 triple-phase CT datasets.

The authors of [27] present a patch-based label propagation that uses relative geodesic distances to define patient-specific coordinate systems as spatial context. This approach is said to overcome the problem of incorrect patch selection, which arises from registration errors after image alignment. Using a dataset of 100 CT image volumes, a mean DSC of 65.5% is achieved with a standard deviation of 18.6%.

As discussed in [25] the proposed approach describes an automatic multi-organ segmentation methodology that is based on spatially-divided probabilistic atlases and performs Markov random field (MRF)-based registration. Experimental results for pancreas segmentation achieve a mean DSC of 69.1% and a standard deviation of 15.3% using 100 CT volumes.

²Image registration is a process that aims to correctly align a target image with a reference image.

³A triple-phase CT scan is a CT technique acquires images at 3 different time points, or phases, following the administration of a contrast.

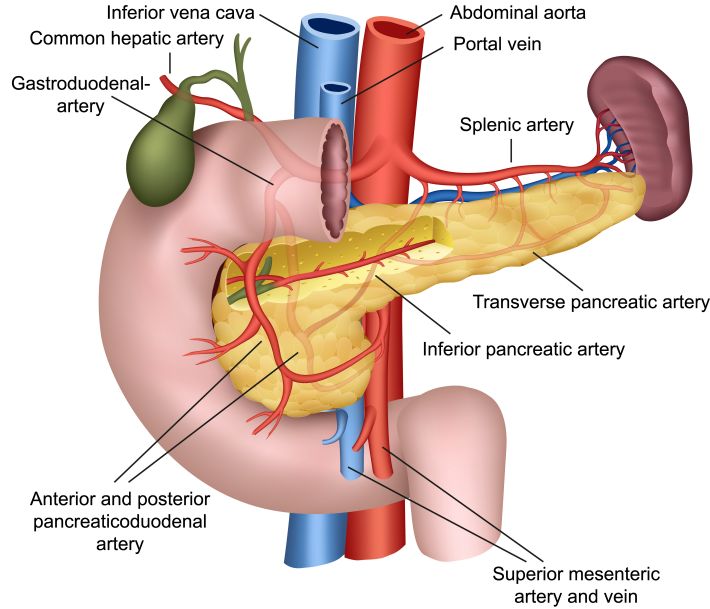


Figure 3.2: Anatomical relationships of the pancreas with surrounding organs and structures. The head of the pancreas lies in the loop of the duodenum as it leaves the stomach; the tail of the pancreas lies near the spleen; and the body of the pancreas lies posterior (nearer the rear) to the outer region of the stomach between the tail [176].

The methodology presented in [21] employs locally weighted atlas selection and a patch-based segmentation scheme to perform multiple abdominal organ segmentation. After a post-processing stage of graph-cuts using a learned intensity model, experimental results achieve a mean DSC of 69.6% and a standard deviation of 16.7% in a dataset containing 150 CT volumes.

The method described in [23] employs voxel-wise local atlas selection, dictionary learning and sparse coding techniques to produce target-specific priors for segmentation. This approach achieves a mean DSC of 71.1% with a standard deviation of 14.7% in a dataset containing 150 CT image volumes.

In the approach reported in [28], statistical shape modelling is performed using conditional shape-location and unsupervised intensity priors. This methodology is applied to abdominal multi-organ segmentation in a dataset containing 134 CT volumes, and for the pancreas alone achieves a mean DSC of 73.4% with a standard deviation of 15.1%. The single-organ segmentation method consists of three main stages including (1) probabilistic atlas-based maximum a posteriori segmentation; (2) statistical shape model-based refinement; and (3) graph-cut refinement.

All of the above methods utilise datasets containing CT image volumes, and although they produce promising segmentation results, these methods suffer from poor statistical stability with high standard deviation ($\geq 14\%$) and require 2 to 4 hours in computation time to process a single image volume [21, 25]. The standard experimental setup in the described atlas-based approaches are performed under the “leave-one-patient-out”

(LOO) cross-validation scheme⁴ for up to $P = 150$ patients or subjects. In a clinical setting, the LOO volume registration from all other patients ($P-1$) as atlas templates (in addition to label fusion) can be computationally inefficient when employing large-scale datasets (containing multiples of 100 image volumes).

The publication described in [70] proposes an atlas selection strategy based on vessel structure around the pancreatic tissue. Pancreas recognition and location is an important building block in the methodology. This method selects atlases with high pancreatic resemblance to the original (unlabelled) image volume. This proposed method employs 150 abdominal contrast-enhanced CT volumes and achieves mean DSC of 78.5% with standard deviation 14.0%.

Although MALP methods have achieved high accuracies ($\geq 70\%$ in mean DSC), inter-patient registration is computationally expensive and extremely poor to imaged organs that possess high variability. Therefore, such methods are best suited for organs that possess relatively lower structural variation in shape and size, such as the liver. For organs, muscular tissue and tumours of high variability, deep learning methods such as convolutional neural networks (CNNs) have rapidly grown in popularity.

3.2.2 Integrating Convolutional Neural Networks

In the last decade, especially since graphics processing units (GPUs)⁵ led the way for a revival of machine learning technologies that were initially introduced in the early 1980s [213], the rise of convolutional neural networks (CNNs) have boosted the performance of many imaging tasks using large-scale data for semantic segmentation [174]. In fact, recent advances in CNNs have been successfully woven into medical image segmentation methods, especially for abdominal organs that are highly deformable and possess vague edge boundaries. Unlike MALP-based algorithms, CNNs do not require selecting a specific atlas nor require deformable registration from training datasets to a target image. A CNN is one of the architectures of deep learning, which in itself is a branch of machine learning. A CNN allows computational models composed of multiple processing layers to ‘learn’ representations of radiological image volumes, and then make a prediction based on that radiological training data in a new, unseen test image volume [37]. Please refer to Appendix B for a brief description about the objective of each layer

⁴“Leave-one-patient-out” (LOO) cross validation: the performance of the algorithm is performed on $P-1$ of the P images and then performance of the modified algorithm is tested on the P -th image. Therefore, in this step, the P -th image is the test image volume and the other $P-1$ images are the training data for optimising the algorithm. Afterwards, this process is repeated P times, each time leaving out a different image to use as the single test image volume. This will result in test performance for all P images.

⁵A Graphics Processing Unit (GPU) is an electronic device designed for handling heavy graphical software applications, well-suited for processing large-scale image data in parallel rather than sequentially.

in a CNN. This ‘learning’ or network training phase optimises the model to predict the location or organ tissue (i.e. greyscale pixel or voxel intensity) in the test image volume. For example, the approach described in [1] describes a bottom-up method by firstly classifying image patches at different resolutions and cascading superpixels. Next, dense image patch labelling is performed using two methods: forest classification on image histogram and texture features; and deep CNN classification on image patches of greater spatial contexts. The proposed method is evaluated on a dataset of 80 CT image volumes and achieves a mean DSC of 70.7%.

The authors of [14] present an approach that uses dense labelling of local superpixel image patches via probability-based CNN and nearest neighbour fusion. Next, a regional CNN samples a set of bounding boxes around each image superpixel at different scales of contexts. The CNN models are trained to assign class probabilities for each superpixel region as either “pancreas” or “non-pancreas”. The methodology proceeds by employing another regional CNN that influences the joint space of CT image pixel intensities and the initial CNN outcome probability maps. Post-processing follows an application of 3D Gaussian smoothing and 2D conditional random fields (CRFs) in order to produce the final segmentation. This updated approach achieves a mean DSC of 71.8% with a standard deviation of 10.7% in a dataset containing 82 CT image volumes.

The use of CNN models is extended in the approach reported in [24] where pancreatic detection and boundary segmentation of MRI volumes is performed using two types of CNN methods. Firstly, a region detection stage separates local image regions as “pancreas” or “non-pancreas” with spatial intensity context; secondly, the boundary detection stage extracts the semantic boundaries of the pancreas. Resulting outcomes from these two networks are merged as the input to a CRF, which provides the final segmentation result. This approach achieves a mean DSC of 76.1% with a standard deviation of 8.7% in a dataset containing 78 MRI volumes.

3.2.3 Baseline Architectures: U-Net, DenseNets and Residual

One of the main motivations of the CNN architecture is optimising the extraction of “useful” features of significance into a vector for an image classification problem. The progression of image classification to segmentation implies having to “classify” every pixel or voxel within an image, which in itself is a computational expensive task compared to classification. The U-Net architecture, which by default is 2D based, attempts to address this challenge, and in doing so has inspired many new architectures in comparison to previous CNN models. In the U-Net architecture, the same features that are used for downsampling (or pooling) can be used to expand a vector of information to the segmented image, and thus preserve the structural and boundary contextual information.

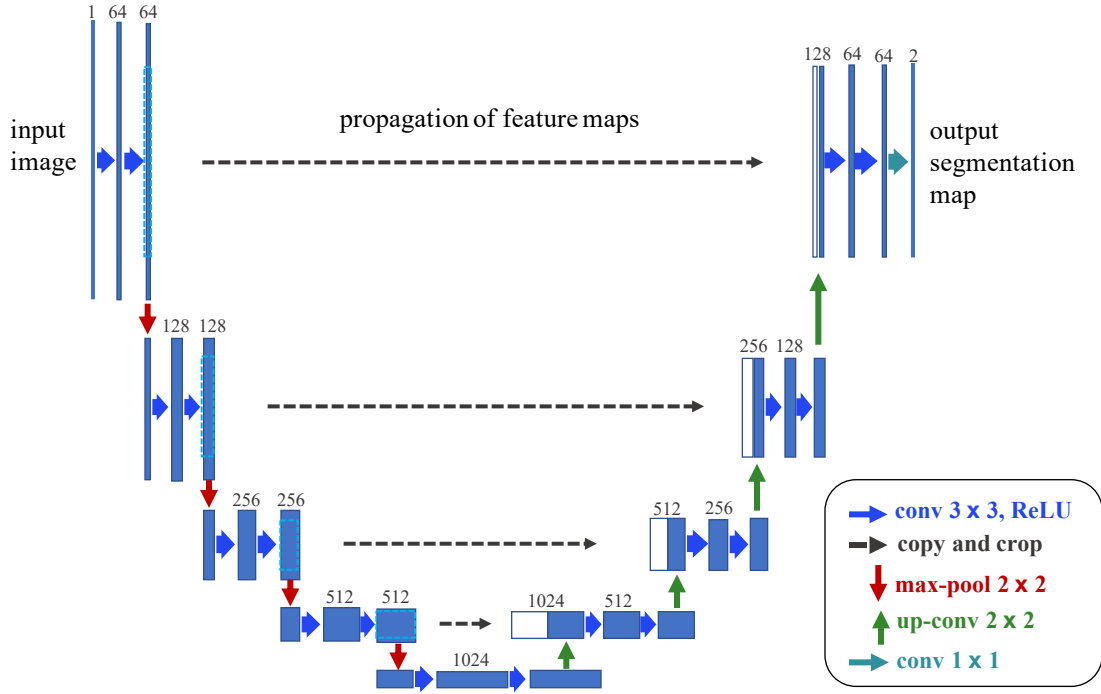


Figure 3.3: U-Net architecture as originally described in [171]. Every blue block represents a multi-channel feature map, in which the number of channels is denoted on top of the block. All white blocks represent copied feature maps and the arrows represent the different operations.

The U-Net architecture [171] consists of three main sections: encoder (downsampling), bottleneck and decoder (upsampling) as illustrated in Figure 3.3, where every blue block represents a multi-channel feature map. In the encoder section, every downsampling step (see red arrows in Figure 3.3) downsamples the previous feature mapping using the output from the previous block. Every level in the encoder receives an input and applies two 3×3 convolutions, each followed by a rectified linear unit (ReLU), and then a 2×2 max pooling. The number of feature maps proceeding each level in the encoder increases twice, enabling the network to recognise different contextual structure and information. The bottleneck layer bridges the encoder and the decoder section through two 3×3 convolutions, each followed by a ReLU.

Next, every level in the decoder section receives an input and applies two 3×3 convolutions, each followed by a ReLU, and then by a 2×2 upsampling (see green arrows in Figure 3.3). In order to maintain symmetry with the encoder section, the number of feature maps after each upsampling step is halved. Furthermore, the input is appended by the cropped feature maps of the corresponding encoder block, ensuring that the features ‘learned’ while encoding the image will be used to reconstruct the image at the segmentation output stage. The number of upsampling steps matches the number of downsampling steps. The final feature mapping is processed through a 1×1 convolution with the number of feature maps equivalent to the defined number of image segments

Layers	Output Size	DenseNet-121	DenseNet-169	DenseNet-201	DenseNet-264
Convolution	112 x 112	7 x 7 conv, stride 2			
Pooling	56 x 56	3 x 3 max pool, stride 2			
Dense Block (1)	56 x 56	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$
Transition Layer (1)	56 x 56	1 x 1 conv			
	28 x 28	2 x 2 average pool, stride 2			
Dense Block (2)	28 x 28	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$
Transition Layer (2)	28 x 28	1 x 1 conv			
	14 x 14	2 x 2 average pool, stride 2			
Dense Block (3)	14 x 14	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 24$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 48$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 64$
Transition Layer (3)	14 x 14	1 x 1 conv			
	7 x 7	2 x 2 average pool, stride 2			
Dense Block (4)	7 x 7	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 16$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 48$
Classification Layer	1 x 1	7 x 7 global average pool			
		1000D fully-connected, softmax			

Figure 3.4: DenseNets architectures for image classification of ImageNet [180]. The growth rate for all the networks is $k = 32$. Each “conv” layer corresponds the sequence of layers: Batch Normalisation, rectified linear units (ReLU) and convolution.

in the output.

Another popular CNN architecture that has inspired recent state-of-the-art approaches is Dense Convolutional Networks (DenseNets) [180]. In this architecture (Figure 3.4) each layer connects to every other layer in a feed-forward manner, and unlike traditional CNNs that have the same number of layers and connections, DenseNets possesses N layers and $N(N + 1)/2$ direct connections. For each layer, the feature-maps of *all* previous layers are used as inputs, and the feature maps produced as outputs are used as inputs into *all* successive layers. Furthermore, concatenating feature maps ‘learned’ by different layers increases variation in the input of successive layers and improves network training optimisation. Since each layer receives feature maps from all preceding layers, the network has a fewer number of channels, enabling higher computational and memory efficiency. The advantages of DenseNets include alleviating the vanishing gradient problem⁶ and significantly reducing the number of parameters.

Residual learning blocks [178] uses a skip connection in which an original input is also added to the output of a convolution block (see third network from the left on Figure 3.5). Consequently, this add-on alleviates the problem of the vanishing gradient by permitting an alternative path for the gradient to flow through. The shortcut connections perform identity mapping (Figure 3.6) and their outputs are added to the outputs of the stacked layers. An entire network can be trained by stochastic gradient descent

⁶The vanishing gradient problem is a fault found in training artificial neural networks with gradient-based learning methods and backpropagation, where each of the neural network’s weights receives an update proportional to the partial derivative of the error function with respect to the current weight in each iteration of training. In some cases during the backpropagation, the gradient will be vanishingly small and thus prevent the weight from changing its value.

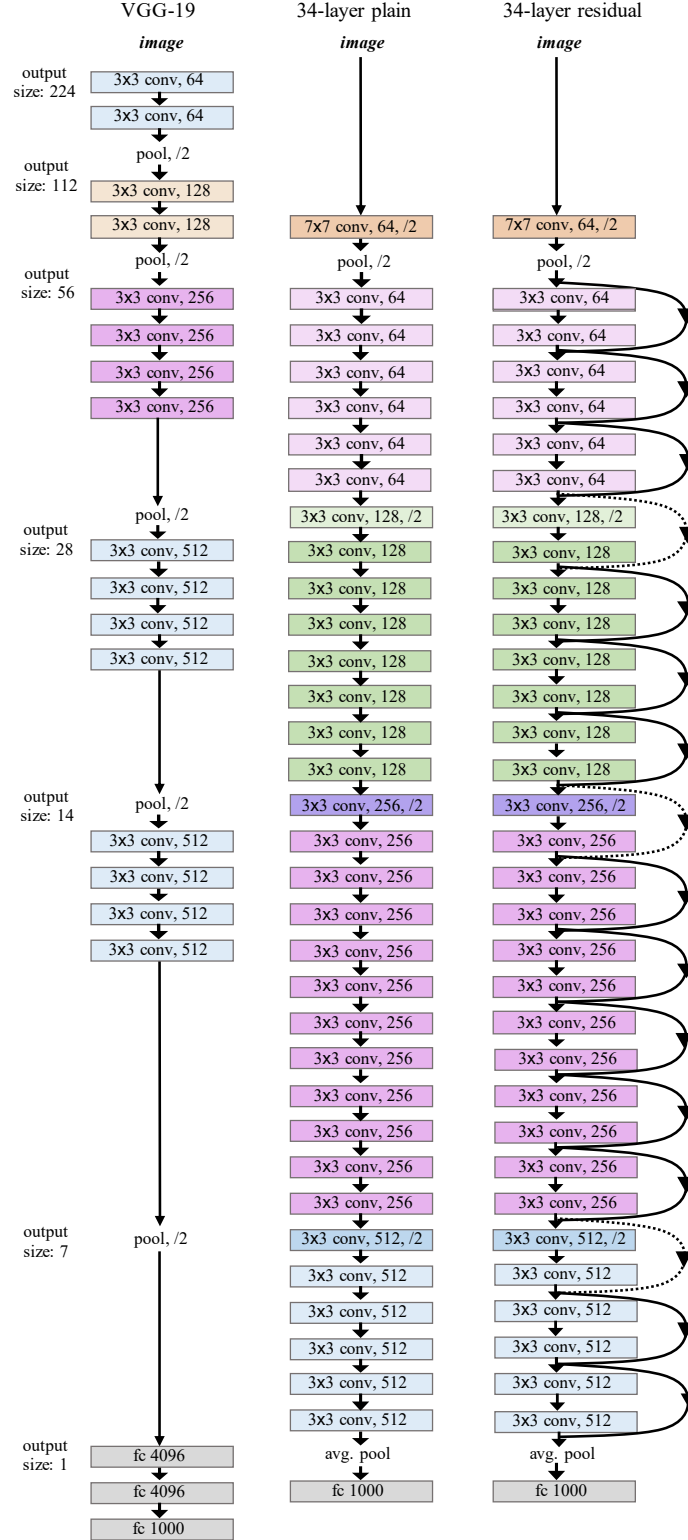


Figure 3.5: The residual architecture implemented for ImageNet [178]. From left to right, the VGG-19 model [97] serves as a reference; a plain network with 34 parameter layers; and a residual network with 34 parameter layers. The dotted shortcuts represent an increase in the dimensions from input to output.

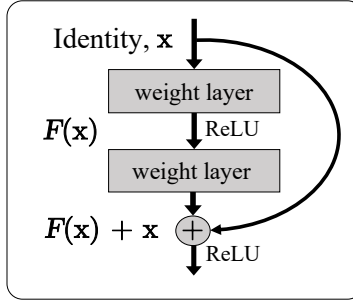


Figure 3.6: A building block for residual learning. Whereas in traditional neural networks, each layer feeds into the next layer, each layer in a network with residual blocks feeds into the next layer and directly into the layers about some “jumps” away [178].

(SGD)⁷ with backpropagation⁸.

3.3 Deep Learning for Medical Image Segmentation

This section delivers a selection of diverse, state-of-the-art deep learning based methods that incorporate key aspects of the baseline architectures as described above, in particular using the encoder-decoder architecture. Section 3.3.1 examines DenseVNet [133] as a method that utilises the advantages of a fully convolutional neural network (FCN)⁹ and 3D volume input deep learning. This method aims to lower the memory and computational costs of the dropout layer¹⁰. Section 3.3.2 proposes the Recurrent Saliency Transformation Network [134], in which a saliency transformation module repeatedly converts the segmentation probability map from the previous iteration as spatial weights in the training stage and applies these weights to the current iteration. This method aims to address the problem of missing contextual information for coarse and fine-stage segmentation. Section 3.3.3 describes a method that reports a reduction in false-positive predictions, using weighted FCNs and multi-atlas spatial atlas priors. Section 3.3.4 explores two different methods that aim to optimise the Jaccard index metric directly through a novel loss function during the training process. Last but certainly not least, Section 3.3.5 highlights various limitations of reported MALP and deep learning-based methods, especially for pancreas segmentation in MRI.

⁷Stochastic gradient descent (SGD) is an iterative method for objective function optimisation.

⁸Backpropagation is an algorithm for supervised learning of artificial neural networks, in which the method updates the artificial neural network’s weights.

⁹A Fully Convolutional Neural Network (FCN) is a variation of a CNN, where the last fully connected layer is substituted by another convolution layer with a large “receptive field”. Therefore, a FCN indicates that the neural network is composed of convolutional layers without any fully-connected layers usually found at the end of a conventional CNN.

¹⁰The dropout layer (using the DropOut method) greatly improves CNN performance, especially in the case where there is limited training data. The DropOut method aims to prevent overfitting; this method reduces the complex interactions between neurons, enabling the ‘learning’ of more ‘defined’ features, which will improve generalisation.

3.3.1 DenseVNet

The publication [133] presents a deep learning segmentation algorithm for eight different organs that are relevant for navigation in endoscopic pancreatic and biliary procedures, including the pancreas. The DenseVNet segmentation network is presented to enable high-resolution activation maps through (memory-efficient) dropout layers and the reuse of feature maps. Two abdominal CT datasets containing a total of 90 image volumes were employed to evaluate this approach, achieving a mean DSC of 78.0% for pancreas segmentation alone. In this approach, a batch-wise spatial dropout scheme is described, which consequently lowers memory and computational costs of dropout. The proposed segmentation method uses a fully convolutional neural network (FCN), in which the architecture design can be understood in terms of 5 key features including batch-wise spatial dropout, dense feature stacks, V-network¹¹ downsampling and upsampling, dilated convolutions¹², and an explicit spatial prior. More specifically, in this network architecture, the DenseVNet has a downsampling sub-network consisting of three dense feature stacks connected by downsampling stride convolutions; each skip connection is a single convolution of the corresponding dense feature stack output, and the up-sampling network consists of bilinear upsampling to the resulting (final) segmentation resolution. The memory efficiencies of dense feature stacks and a proposed batch-wise spatial dropout allow deeper networks at higher resolutions, which is advantageous for segmentation of smaller structures such as the pancreas.

The loss function that is used for training DenseVNet is the weighted sum of a L2 regularisation loss with label-smoothed probabilistic Dice similarity coefficient (DSC) scores for each organ (denoted as x) averaged across subjects in each mini-batch, as described below:

$$pDice_x(L''_x, R_x) = \left(\frac{\min(L''_x, 0.9) \cdot R_x}{\|R_x\|_2 + \|\min(L''_x, 0.9)\|_2} \right)$$

where vectors $L''_x = softmax(L'_x)$ and R_x are the algorithm's probabilistic segmentation and the binary reference standard segmentation for organ x for each subject, respectively. In order to further reduce the extreme class imbalance, DSC scores that hinge losses heavily penalising DSC scores below 0.01 and 0.10 were introduced after periods of 25 and 100 iterations, respectively. Thus, the loss function at iteration i can be described as below:

¹¹A V-network architecture employs 3D input volumes, and comprises of downsampling and up-sampling subnetworks with skip connections to propagate higher resolution information to the final segmentation.

¹²Dilated convolution is the process of applying a filter to a larger receptive field compared to standard convolution.

$$loss(L'', i) = \sum_{\forall W} \frac{\overline{w^2}}{40} - \frac{1}{8} \sum_{x=1}^8 d(pDice(L''_x, R_x), i)$$

$$d(l, i) = l + 100h(l, i, 0.01, 25) + 10h(l, i, 0.1, 100)$$

and

$$h(l, i, v, t) = \text{sigmoid}\left(\frac{6(i-t)}{t}\right) \left(\frac{\max(0, v-l)}{v}\right)^4$$

where $w \in W$ are filter values, l is the DSC loss, v is the hinge loss threshold, and t is the delay in iterations.

For computational and memory efficiency, the new batch-wise spatial dropout is introduced by modifying the convolution filters denoted by:

$$\bar{X}_i = c(\bar{X}_{i-1}, \bar{o}(W, B^I, B^O), s, \bar{y})$$

where c is a convolutional unit, W is a convolutional filter, B^O is a stochastic binary mask, B^I is a dropout mask, and $\bar{o}(W, B^I, B^O)$ is a new filter without input and output channels masked by B^I and B^O . \bar{X}_{i-1} is the output of the previous network layer after batch-wise spatial dropout, and s is the stride and \bar{y} is the scale parameter of non-dropped channels. Note that \bar{X}_i is identical to the non-dropped channels of \hat{X}_i but does not compute or store the dropped channels, and that subsequent convolutions are unaffected if their filter is similarly modified. The proposed batch-wise spatial dropout for regularisation does not use 1×1 bottleneck layers. Specifically, the output of an m -layer dense feature stack $f_m(X_0) = ([X_0; X_1; \dots; X_m])$ where, as described below,

$$X_i = \hat{f}_i([X_0; X_1; \dots; X_{i-1}])$$

and

$$\hat{f}_i(X) = c(X, \bar{o}(W_{a,n_f,d_i}, B_i^I, B_i^O), 1, \bar{y})$$

where $[A; B]$ denotes channel-wise concatenation; W_{a,n_f,d_i} is an $a \times a \times a$ convolution filter with n_f output channels, and dilation rate, d_i . Furthermore, $B_i^I = [B_0^O; B_1^O; \dots; B_{i-1}^O]$ selects all previously computed channels, B_0^O selects all channels from X_0 and otherwise B_i^O is sampled stochastically, such that $\lceil pn_f \rceil$ channels are selected ($p = 0.5$ and n_f scales the convolutional unit outputs). In this proposed scheme, the feature stacks inherently encode identity functions as the final output channels include the inputs; secondly, the feature stacks combine multiple network depths within a single network,

allowing effective propagation of gradients through the network; and third, when memory constraints limit the number of activation maps, information from earlier layers is stored only once in memory, but can be accessed by successive layers.

3.3.2 Recurrent Saliency Transformation Network

The approach reported in [134] proposes an intuitive approach to finding pancreatic ductal adenocarcinoma (PDAC), the most common type of pancreatic cancer, by examining abdominal CT scans. A two-stage framework is proposed: the first stage which segments the pancreas into a binary mask, and the second stage compresses (or restructures) the mask into a shape vector and performs abnormality classification. In this approach, shape representation and classification are performed to exploit expert knowledge that PDAC often changes the shape of the pancreas, and to prevent overfitting. Experiments are evaluated on a dataset containing 136 (diagnosed with PDAC) and achieve mean DSC of 71.45%. The overall framework in this approach can be founded as follows. A CT image volume, \mathbf{X} , is a $W \times H \times L$ matrix, where W , H and D are the width, height and length of the cube, respectively. Each element in the cube represents the Hounsfield unit (HU) at the specified position. Each volume is annotated with a binary pancreas mask \mathbf{S}^* which shares the same dimensionality with \mathbf{X} . A discriminative function is described as $p(\mathbf{X}) \in \{0, 1\}$, with 1 indicating that this person suffers PDAC and 0 otherwise. The goal is to decompose this discriminative function into two stages: the first stage is a segmentation model $\mathbf{f}(\cdot)$ for voxel-wise pancreas segmentation, where $\mathbf{S} = \mathbf{f}(\mathbf{X})$; and the second stage is a mask classifier $c(\cdot)$ which assigns a binary label to the mask, \mathbf{S} . In order to incorporate the shape information, $c(\cdot)$ is further decomposed into a shape encoder, $\mathbf{g}(\cdot)$ which produces (1) vector $\mathbf{v} = \mathbf{g}(\mathbf{S})$ to depict the shape properties of the binary mask \mathbf{S} , and (2) a shape classifier $h(\cdot)$ which predicts whether the shape vector \mathbf{v} corresponds to a PDAC-pancreas. Therefore, the overall framework can be described as below:

$$p(X) = c \circ \mathbf{f}(\mathbf{X}) = h \circ \mathbf{g} \circ \mathbf{f}(\mathbf{X})$$

The pancreas segmentation approach starts with an encoder-decoder network known as 2D recurrent saliency transformation network¹³ (STN), for which it trains three models from the coronal, sagittal and axial planes. Based on a pancreas segmentation function, $\mathbf{S} = \mathbf{f}(\mathbf{X})$, it is important to determine the classification of abnormality of this pancreas, which is achieved by first compressing the segmentation mask into a low-dimensional vector, $\mathbf{v} = \mathbf{g}(\mathbf{S})$, and then applying a classifier $h(\cdot)$ on top of \mathbf{v} . The shape representation network $g(\cdot)$ includes the gradual down-sampling of the seg-

¹³A saliency transformation module repeatedly converts the segmentation probability map from the previous iteration as spatial weights and applies these weights to the current iteration.

mentation mask. Following generative and discriminative voxel modelling with CNNs [135], this network is implemented by a series of 3D convolutional layers. Regarding the dimensionality of the shape vectors (i.e., the number of output neurons), a high-dimensional representation carries more information, but also risks over-fitting under limited training data. The segmentation and shape representation networks perform image down-sampling, with the segmentation network starting with the raw input image and is therefore computationally expensive. The shape representation network is much simpler in comparison with the segmentation network, which processes the entire volume at once. Consequently, this allows the network to be optimised, together with the classifier. In the final step, a $h(\cdot)$ with 2-layer fully-connected network is implemented. The simplicity of $h(\cdot)$ means that the vector \mathbf{v} carries discriminative shape information, which is easy to classify. Being a differentiable module, it can be optimised with a joint shape representation network, which brings about accuracy gain that is consistent.

3.3.3 Multi-atlas Combined FCN Approach

The authors of [136] extend organ segmentation in MRI to present a methodology that combines an atlas-based approach with CNN implementation. This approach builds on previous work described in [30, 139, 140] by incorporating (a) weighting schemes to support class imbalance; and (b) a specialised organ region-of-interest (ROI) selection. Later, spatial information from atlases and CNNs are optimised jointly and applied for organ segmentation. A dataset containing 48 whole MRI body volumes is evaluated to achieve a mean DSC of 61.2% for the pancreas alone. In this approach, a coarse-scale stage is firstly employed to locate the organ of interest or ROI for a subsequent fine-scale organ segmentation; this stage effectively reduces the complexity of the background while enhancing the discriminative information of a small organ (such as the pancreas). Later, at the fine-scale segmentation stage, the ROI selection from the previous stage is considered for coarse organ segmentation. On each stage, two different networks are trained respectively. Similar to the training stage, the testing stage follows an initial coarse-scale network to obtain the rough position of the small organ of interest; and later, fine-scale networks are employed for binary segmentation of this small organ. In order to handle the problem of class imbalance, a weighted-FCN is utilised for both coarse-scale and fine-scale segmentations. In the former segmentation, a weighted-FCN performs multi-class segmentation with different weights set for each class, where relatively large weights are applied for a smaller organ (compared to bigger organs such as the liver), according to the relative size to the whole body volumes. In this approach, ROI selection is important in order to locate the organ of interest in a specific region. Following organ localisation from the coarse-scale segmentation, experiments showed

that in a number of cases, the FCN fails to locate small organs, leading to a much wider ROI for the fine-scale segmentation. Therefore, it is essential to incorporate context and spatial information, including spatial priors in order to guide the second stage network with a better ROI.

Multi-atlas Technique for Producing Spatial Priors

In addition to the weighted-FCN in [136], a multi-atlas technique was chosen for producing spatial priors, because although multi-atlas segmentation accuracy is lower than CNN-based accuracy, this technique generates good organ localisation. In this approach, each atlas image is registered to the target image and then fused; the patch-based segmentation with augmented features method is chosen for the label fusion technique. Unlike [137] that uses a multi-atlas approach for organ localisation at first step before binary segmentation via CNNs, this approach incorporates it (only) after the coarse-scale segmentation. Here, the probability maps of coarse-segmentation are multiplied with the one from the multi-atlas approach to produce the 3D bounding box for fine-scale segmentation.

3.3.4 Optimising the Jaccard Index Directly

The approach in [30] describes a convolutional-recurrent neural network that performs segmentation on image volumes by incorporating neighbouring slice segmentation predictions. Furthermore, a loss function is developed and additional neural networks are trained to optimise the Jaccard index directly. Using a dataset containing 82 CT image volumes, a mean DSC of 82.4% with standard deviation 6.7% is achieved; a second dataset containing 79 MRI volumes is evaluated, achieving a mean DSC of 80.5% with standard deviation 6.7%.

The publication reported in [130] initially trains a 2D CNN sub-network with deep-supervision and multi-scale feature map aggregation, so that it can be trained from scratch with small-sized training data. Afterwards, the successive CNN outputs are processed by a recurrent neural network (RNN) sub-network, which refines the stability of segmented objects in a given image (slice). The RNN sub-network contains convolutional long short-term memory (CLSTM), thus regularising the segmentation of an image by incorporating predictions of its neighbouring slices. The arranged CNN-RNN model is trained end-to-end and evaluated on both MRI and CT image volumes. Using a CT dataset containing 82 image volumes, this approach achieves a mean DSC of 83.3% with standard deviation 5.6%, and achieves mean DSC of 80.7% with standard deviation 7.4% using a dataset containing 79 MRI volumes.

The long short-term memory (LSTM) strategy treats the context information of the pancreas as a sequence, and this context information is assigned to different nodes in LSTM, which can optimise segmentation results with spatial information. The disadvantage of this method is that the variable curvature of the pancreas shape may lead to incorrect guidance; for example, the segmentation result can vary drastically if the head of the pancreas is used to guide the tail of the pancreas. Furthermore, an LSTM-based model serves as a connected but separate module behind other segmentation networks, which has little contribution to the result but adds to the computational costs of the entire network and increases the overall time for that network to converge.

In both [30] and [130], a loss function known as Jaccard (JAC) loss is proposed to train the CNN image segmentation model. This error function aims to optimise the Jaccard index directly during network training. The approximated definition of JAC loss is defined below:

$$\begin{aligned}\tilde{L}_{jac} &= 1 - \frac{\sum_{f \in Y_+} \min(1, \hat{y}_f)}{|Y_+| + \sum_{b \in Y_-} \max(0, \hat{y}_b)} \\ &= 1 - \frac{\sum_{f \in Y_+} \hat{y}_f}{|Y_+| + \sum_{b \in Y_-} \hat{y}_b}\end{aligned}$$

where Y_+ and Y_- is defined as the foreground and background pixel set, respectively, and $|Y_+|$ is the cardinality of Y_+ . From here, \hat{y}_f and $\hat{y}_b \in \{0, 1\}$ are indexed pixel values in Y_+ and Y_- . Theoretically, L_{jac} and \tilde{L}_{jac} share the same optimal solution for predicting a segmentation outcome. The model is then updated by gradient flows as described below:

$$\frac{\partial L_{jac}}{\partial \hat{y}_j} = \begin{cases} -\frac{1}{|Y_+| + \sum_{b \in Y_-} \hat{y}_b}, & \text{for } j \in Y_+ \\ \frac{\sum_{f \in Y_+} \hat{y}_f}{\left(|Y_+| + \sum_{b \in Y_-} \hat{y}_b\right)^2} & \text{for } j \in Y_- \end{cases}$$

where $\hat{y}_j \in \{0, 1\}$ is indexed pixel value in the network's prediction outcome, \hat{Y} . It should be noted that \hat{y}_j is the probability number in the range of $[0, 1]$. In knowledge that the inequality $\sum_{f \in Y_+} \hat{y}_f < \left(|Y_+| + \sum_{b \in Y_-} \hat{y}_b\right)$ is true by definition, the JACLoss assigns larger gradients to foreground pixels that inherently equalises the foreground and background classes. This loss function has reported to empirically work better than the cross-entropy loss when segmenting small objects (e.g., the pancreas) in MRI and CT volumes.

3.3.5 Limitations

In recent years, multi-atlas or MALP methods have achieved high accuracies ($\geq 70\%$ in mean DSC). However, inter-patient registration is computationally expensive and extremely poor on imaged organs that possess high variability such as the pancreas. With the exception of [28] and [136], the CNN methods described in the above yield DSC results that outperform the reported MALP approaches and produce a lower standard deviation (i.e. exhibit statistical stability). However, such convolutional neural networks are prone to suffer from an imbalance between classes (e.g., “pancreas” versus “non-pancreas”) and overfitting during the network training stage [31], and thus ignore features that are related to the organ of interest during the testing stage. Furthermore, the majority of segmentation methods have been performed on CT modality as opposed to the additional difficulties of image artefacts and higher blurred boundaries between organs in MRI.

3.4 Original Contributions to Knowledge

Considering the limitations and challenges of automated medical image segmentation in MRI, this section provides an overview of key technical methodological novelties, covering the contributions to the field of automated object segmentation in medical image analysis. The following list presents the technical and methodological novelties:

- (a) a learned intensity model for digital contrast enhancement in image volumes;
- (b) a hybrid energy-minimisation segmentation approach exploits edge detection to yield detailed, optimal contouring of the pancreas;
- (c) a post-processing stage integrates principal geometric descriptors that characterises tissue and employs radiological expert-knowledge for refined tissue classification;
- (d) a deep learning novel loss (error) function based on the Hausdorff metric and a sinusoidal component.

3.4.1 (a) Digital Contrast Enhancement

Although multi-atlas based approaches [21, 23, 28] have achieved high quantitative accuracy scores ($> 90\%$) on organs such as the liver and kidneys, segmentation of pancreas is far lower in quantitative accuracy. Not only does the pancreas account for less than 1% in a given scan, the selection of atlases with high similarity in the pancreas region is not successful because surrounding tissue, such as the duodenum, have similar greyscale intensity. In Method 1, described in Chapter 4 (Section 4.2), a pre-processing

stage improves the distributional characterisation of intensities between pancreatic and surrounding tissue boundaries. A digital contrast intensity model is proposed to increase greyscale differences in an image volume, particularly in pancreatic contouring. Optimum parameters for tissue enhancement are achieved through analysing intensities in a corresponding training imaging dataset. The motivation is to differentiate the nearby splenic vein, superior mesenteric artery and surrounding tissue that are often overlapping, or in close proximity to the pancreas. Consequently, this stage reduces the number of “non-pancreas” image pixel predictions in later processing.

3.4.2 (b) Hybrid Energy-minimising Segmentation

A 3D based algorithm is employed at a stage in the pancreas segmentation process, as described in Chapters 4 and 5. Consequently, this produces greater consistency in spatial smoothness and prediction among successive slices. In order to further overcome problems caused by the low contrast between organ boundaries, a pre-trained edge detection model is utilised to “strengthen” the boundaries of distinct pancreas contours and reduce edges that appear within these closed contours, consequently improving tissue classification at a post-processing stage. Thus, the major pancreas region and surrounding tissue are extracted using a hybrid segmentation approach that integrates random forest probability-wise predictions of superpixels; structured forest edge detection for local regions; and energy-minimising 3D continuous max-flow and min-cuts, together which produces very detailed boundary preservation. Section 4.3 (Chapter 4) further details this proposed stage.

3.4.3 (c) Multi-level Geometrical Descriptor Analysis and Tissue Classification

Unlike previous approaches [1, 14, 23, 27] that label fixed image patches as “pancreas” or “non-pancreas” towards or during the final methodology stage, the approach presented in Chapter 4 demonstrates the effectiveness of classifying distinct contours with diverse size and structure. A multi-level stage of post-processing is proposed for refined tissue classification: morphological operations are performed to obtain geometrical descriptors in an image volume and positioning of distinct contours are analysed to determine the likelihood of being pancreatic tissue. A combinations-based method of elimination aims to remove non-pancreatic contours that are likely to represent a fraction of the inferior vena cava, common hepatic artery, splenic artery and vein, superior mesenteric, stomach, left kidney or spleen. This detail of analysis, which combines expert-knowledge of radiology and anatomical structure, preserves the original contouring of the pancreas while eliminating surrounding tissue separated by a small pixel range of [1, 5].

3.4.4 (d) Loss Function

The approach proposed in Chapter 5 incorporates a deep learning network that is trained according to a novel loss or error function. The motivation for this new loss function is to (a) accelerate the training (reduce computational costs), and (b) enhance final segmentation results in comparison to the expert-led ground-truth (annotation) of the pancreas organ. Empirically, this function performs better than the conventional cross-entropy loss [68] when segmenting the pancreas in CT and MRI volumes. Section 5.2.2 (Chapter 5) mathematically describes this novel loss function and Section 6.4.2 (Chapter 6) discusses and demonstrates its effective performance against cross-entropy loss.

3.4.5 Attributes of Contributions

Boundary Preservation and Contouring

The final quantitative pancreas segmentation results are better than or close to state-of-the-art approaches [1, 14, 21, 23, 24, 25, 27, 28, 30, 32, 33] for both CT and MRI modality, and report higher statistical stability with lower standard deviation metrics. Furthermore, the feedback received by an expert radiographer and an expert radiologist, reveals that the proposed method produces detailed contouring of the pancreas for every protrusion and indentation as opposed to an approximate or mean tracing of the organ.

Multimodality

The proposed methods in Chapters 4 and 5 deliver segmentation accuracy results with comparable consistency and robust statistical stability across multiple modalities and datasets. In contrast to several reported strategies that employ only CT [1, 14, 21, 23, 25, 27, 28, 32, 33] or two modalities but one dataset [24, 30], the proposed methods are evaluated on two imaging modalities including a CT dataset, and two MRI datasets that were obtained using two different MRI machines and two different scanner imaging protocols.

Robustness to Noise Variation

The proposed methods in Chapters 4 and 5 are shown to be effective across image volumes with varying noise and distortion. Unlike other publications that focus on one modality or one dataset for every modality, this method is robust to variations in noise, distortion, sharpness and changes in greyscale intensity distribution within multiple MRI datasets and a CT dataset. Section 7.2.2 (Chapter 7) provides further detail and includes three different visual examples.

Computational Efficiency

Unlike previous non-deep learning-based publications that require 2 to 4 hours to evaluate a single image volume [21, 25] the approach described in Chapter 4 evaluates an image volume within 10 to 35 minutes via a workstation with i7-5930k-CPU at 3.50 GHz. This run-time can be potentially reduced by a factor of 10 by using a GeForce Titan X GPU. Furthermore, the improved deep learning approach of Method 2, that is described in Chapter 5, trains the main network in significantly less time of approximately 11 hours versus approximate times of state-of-the-arts: 2 - 3 days [14, 173], 1 day [171, 174] and 21 hours [133].

Generalisability

The proposed methodologies in Chapters 4 and 5 can be further optimised and applicable to other abdominal MRI and CT sequences and also, generalisable to other organ, tumour or muscular tissue segmentation tasks. For every organ of interest, it is essential to incorporate corresponding ground-truth annotation and expert radiological knowledge about the organ’s shape, size, curvature and location. Chapter 8 presents and discusses the significance for a generalised segmentation framework and details exemplar results for MRI liver, iliopsoas muscles and kidneys segmentation.

3.5 Conclusion

Automatic segmentation of the pancreas remains challenging compared to the automatic segmentation of other major organs such as the heart, liver, kidneys or spleen. The high structural variability and location of the pancreas creates further challenges for accurate extraction in medical image scans. Nevertheless, pancreas segmentation is a prerequisite for computer-aided diagnosis systems (CADx) in order to support the identification of tumour differences, determine the progression of type 2 diabetes, or provide a “second opinion” that can help to reduce possible misdiagnosis. This chapter addresses the challenges and motivation for automatic pancreas segmentation and visualisation in 3D radiological scans of multiple modalities. Furthermore, this chapter presents an overview of recent work developed in this field of scope and also, the contributions of novel approaches for pancreas segmentation that will be further detailed in proceeding chapters.

Chapter 4

Method 1: Morphological and Multi-level Geometrical Descriptor Analysis

4.1 Introduction

This chapter presents a novel approach (Method 1) for automatic pancreas segmentation in Magnetic resonance imaging (MRI) and Computer tomography (CT) scans. This method exploits 3D segmentation that, when coupled with geometrical and morphological characteristics of abdominal tissue, classifies distinct contours in close pixel-range proximity as “pancreas” or “non-pancreas”. The term ‘contour’ in this context refers to a distinct binarised shape or segment that is extracted from the image volume during the automatic segmentation process. Although “pancreas” or “pancreas tissue” refers to the medical terminology representing the grouping of pancreas’ cells, these terms describe the contour that is classified by the algorithm in the proposed method to represent actual, pancreas tissue. There are three main stages of this approach: (1) identify a major pancreas region and apply contrast enhancement to differentiate between pancreatic and surrounding tissue; (2) perform 3D segmentation by employing continuous max-flow and min-cuts approach, structured forest edge detection, and a training dataset of annotated pancreata; (3) eliminate non-pancreatic contours from resultant segmentation via morphological operations on area, structure and connectivity between distinct contours. This section provides an overview of the first proposed approach (Method 1), accompanied by an illustration for additional guidance. Section 4.2 discusses the motivation and describes an intensity model that enhances the distinction between pancreatic tissue and the surrounding tissue in a 3D radiological scan (image volume). This section also addresses the need for isolating a major pancreas region that discards the liver and major sections of the kidneys. Section 4.3 presents the primary segmentation

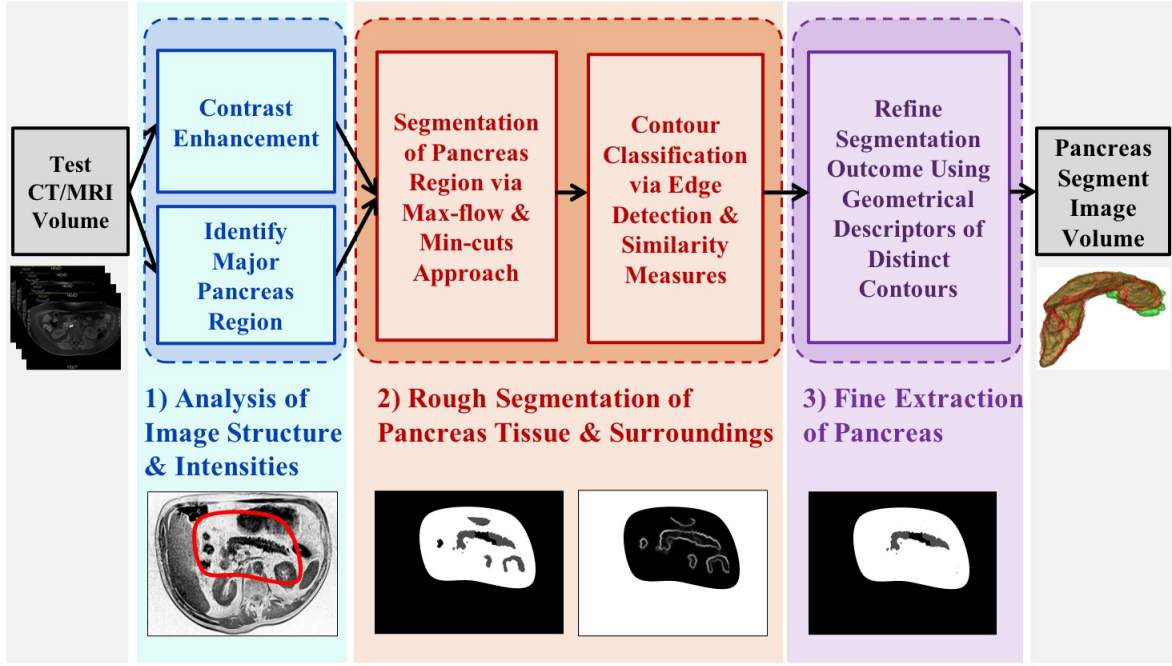


Figure 4.1: Overview of methodology: three main stages are proposed for automatic pancreas segmentation in abdominal CT and MRI scans (volumes).

approach that produces distinct tissue for further classification. Section 4.4 discusses three main levels of refinement that progressively target segmented tissue that lies in close proximity to the pancreas but are identified as false “pancreas” predictions and discarded from the final pancreas segmentation. Section 4.5 concludes with a summary of this chapter.

The methodology of the proposed approach, as illustrated in Figure 4.1, progresses through three main stages, each one of which is discussed in proceeding sections. This approach follows a sequential process of applying a digital contrast enhancement to an image volume, which aims to improve intensity differentiation between pancreatic and non-pancreatic tissue. An artificial neural network is used to achieve optimum parameters for enhancement. Next, a major pancreas region is identified using a trained random forest and probability-wise predictions of superpixels in each 2D image (slice) within the image volume. An initial 3D segmentation is performed via continuous max-flow and min-cuts approach, and structured forest edge detection. Afterwards, a stage of post-processing eliminates remaining non-pancreatic tissue via morphological operations on area, curvature, position and gradient between distinct contours in the segmented image volume.

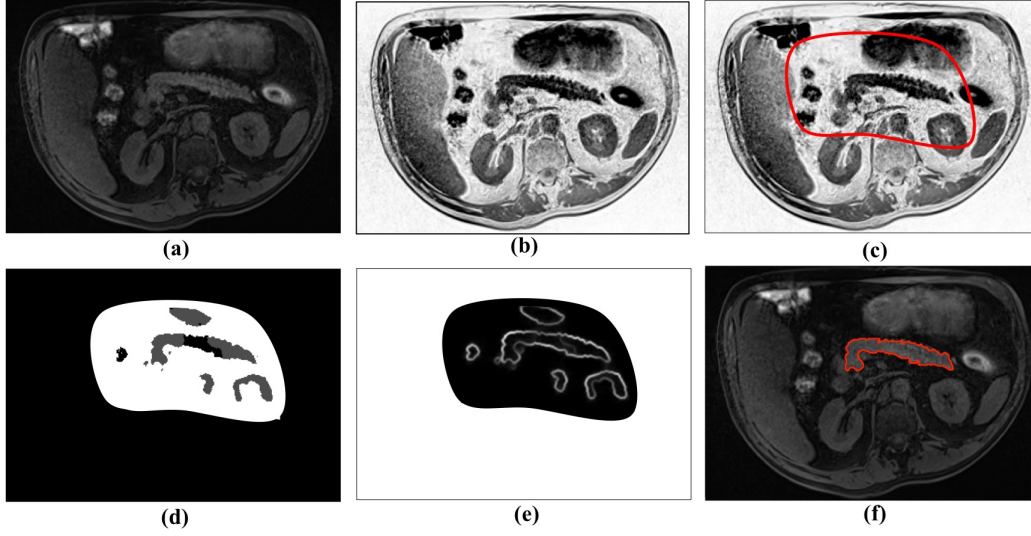


Figure 4.2: Visualisation of results for a slice from image volume. (a) Original MRI slice, (b) MRI slice after contrast enhancement, (c) Closed red outline encapsulates major pancreas region, (d) Segmentation following max-flow and min-cuts approach, (e) Boundary detection using structured forest learning, (f) Final contour segmentation after refinement.

4.2 Tissue Enhancement and Elimination

An effective application of contrast enhancement in CT or MRI can differentiate pancreatic tissue and boundaries against background classes of blood vessels, stomach fundus and the first section of the small intestine (duodenum). Figure 4.2(a)(b) depicts a slice from an image volume before and after contrast enhancement, respectively. Next, a major pancreas region is identified using a random forest based method reported in [1]. Figure 4.2(c) displays a red outline over a sample slice that represents the predicted major pancreas region for that particular slice in an image volume.

4.2.1 Digital Contrast Enhancement

A sigmoid function is applied to a test image volume by incorporating a gain, g , which controls the actual contrast, and a cut-off value, c , which represents the normalised greyscale value about which contrast level is changed. Every i -th slice, s_i , in the image volume undergoes contrast enhancement, $C(s_i)$, as described in Equation 4.1:

$$C(s_i) = \frac{1}{1 + \exp [g(c - s_i)]} \quad (4.1)$$

Empirically tested, the values for gain and cut-off were assigned to image volumes in the training dataset by taking into consideration their respective mean greyscale intensity

value. From here, an artificial neural network¹ (consisting of two hidden layers, the former with 20 and the latter with 10 hidden neurons) is developed to predict the gain and cut-off value for a test image volume. The network training is optimised using the “mean squared error” loss over 100 epochs. In general, the value of the gain and cut-off increases with respect to an increasing mean intensity value of the whole image volume.

4.2.2 Identify Major Region of Interest

Next, a major pancreas region is identified using the method reported in [1]. A random forest is trained on a selection of extracted features in the training data; afterwards, the trained forest predicts the likelihood of a region in the test image volume as “pancreas” given a probability threshold. To elaborate, every image volume in the training dataset is initially converted to superpixels [104] of 32 pixel region size. From here, feature information is extracted from image patches of 25×25 pixels, describing texture and the probability of a patch as “pancreas” based on voxel intensity analysis in the training data and in accordance to corresponding ground-truth pancreas annotation. Equation 4.2 describes this probability as:

$$p^+ = \frac{f^+(M^+)}{(f^+(M^+) + f^-(M^-))} \quad (4.2)$$

where $f^+(M^+)$ and $f^-(M^-)$ are filter density estimators representing the intensity distributions of the positive $\{M^+\}$ and negative $\{M^-\}$ “pancreas” and “non-pancreas” voxels from an image volume, respectively.

Two further features for each axial slice are the normalised relative x -axis and y -axis positions in the range $[0, 1]$, which are calculated at each image patch centre. From here, a total of 46 image patch-level features per superpixel are employed to train the random forest classifier. The image patch labels are acquired by extracting information from their patch centre pixels, again, in accordance to their ground-truth pancreas annotation. The trained random forest classifier distinguishes the “non-pancreas” class patches such as the liver and greater outer region. Figure 4.2(c) displays a red outline over a sample slice that embodies the area predicting “pancreas” at a probability threshold (0.85) per superpixel - this represents the major pancreas region for that particular slice in a test image volume.

¹Artificial neural network consists of elements called artificial neurons, which receive input(s), change their hidden state (or activation) according to those input(s), and produce output(s) depending on the input(s) and activation(s).

4.3 Rough Segmentation of Pancreas

Following tissue enhancement and elimination, the identified major pancreas region of interest as a volume is processed through an unsupervised 3D segmentation algorithm described in [2], which uses maximal-flow and minimum graph-cuts approach constrained by predefined key parameters that are described in Chapter 6, Section 6.4.1. Figure 4.2(d) displays the resulting segmentation of a single slice in a test image volume.

The basis of this algorithm relies on graph theory, where an image is represented as a graph; each pixel is viewed as a node in the graph and edges are formed between nodes. Nodes are interconnected to neighbouring nodes. A graph can be represented as $G = \{V, E\}$ where V is the set of vertices or nodes and E is the set of edges, and $E \subset V \times V$. The vertex set of nodes are positioned in a 2D or 3D nested grid, linked with two terminal nodes of source, s , and sink, t , which do not correspond to any pixels directly. The set of edges can be divided into two groups: the terminal edge, e_s or e_t connects the corresponding terminal node to each grid node respectively; the spatial edge, e_i , follows the path of the given grid and connects two neighbouring grid stencils (with the exception of s and t). A non-negative cost or loss denoted by $L(e)$ is assigned to each edge, e . The algorithm discussed as follows concerns graphs represented in 2D domain denoted by Ω but this can be extended to a 3D context. A graph-cut partitions the spatial nodes of Ω into two disjoint segment sets, one concerns the source, s , and the other to the sink, t , as described below:

$$V = V_s \cup V_t \quad (4.3)$$

and

$$V_s \cap V_t = \emptyset \quad (4.4)$$

The total cost or loss of edges $e \in E_{st} \subset E$ can also be described as the energy of each graph-cut, whose terminal points are assigned to two different and distinct partitions. The minimum graph-cut approach aims to find the two partitions of vertices where the corresponding graph-cut energy is minimised, such that:

$$\min_{E_{st} \subset E} \sum_{e \in E_{st}} L(e) \quad (4.5)$$

If each edge, $e \in E$ is represented as a tube-like channel, then the edge loss $L(e)$, can represent the maximum limit of the tube restricted by the following conditions described below.

For an undirected grid edge, $e_i \in E$, the spatial flow, $p(e_i)$ is limited to:

$$|p(e_i)| \leq L(e_i) \quad (4.6)$$

If a source edge, $e_s(v)$, connects s to $v \in V \setminus \{s, t\}$, then the source flow, $p_s(v)$ is directed from s to v . The maximum limit, $L_s(v)$ shows that:

$$0 \leq p_s(v) \leq L_s(v) \quad (4.7)$$

For a sink edge, $e_t(v)$ that connects $v \in V \setminus \{s, t\}$ to t , the sink flow $p_t(v)$ is directed from v to t . The maximum limit, $L_t(v)$, shows that:

$$0 \leq p_t(v) \leq L_t(v) \quad (4.8)$$

If $I(v) \subset E$ is the set of edges connecting $v \in V \setminus \{s, t\}$ to its neighbouring nodes, then the energy flows passing v are constrained by:

$$\left(\sum_{e_i \in I(v)} p(e_i) \right) = p_s(v) - p_t(v) \quad (4.9)$$

By taking into account the conditions of 4.6, 4.7, 4.8 and 4.9, the maximal flow approach aims to find the greatest amount of energy allowed to flow from the source, s , so that:

$$\max_{p_s} \sum_{v \in V \setminus \{s, t\}} p_s(v) \quad (4.10)$$

Indeed, it has been shown that the maximal-flow problem in 4.10 is equivalent to that of the minimum-cut problem described in 4.5.

The maximal flow approach is now extended to a continuous domain. Thus, Ω can be used to represent a closed and *continuous* 2D or 3D domain with terminals source, s and sink, t . At every position, $x \in \Omega$, the spatial flow passing x can be written as $q(x)$. Additionally, the directed source flow from s to x can be denoted by $q_s(x)$ and the directed sink flow from x to t by $q_t(x)$. The continuous max-flow approach can be developed in the same manner as discussed above for discrete max-flow. Taking into consideration 4.6, 4.7, 4.8 and 4.9 over graph G , the flow functions $q(x)$, $q_s(x)$ and $q_t(x)$ can have similar constraints imposed over the spatial domain, Ω , such that:

$$|q(x)| \leq C(x) \quad (4.11)$$

$$q_s(x) \leq C_s(x) \quad (4.12)$$

$$q_t(x) \leq C_t(x) \quad (4.13)$$

$$\nabla \cdot q(x) = q_s(x) - q_t(x) \quad (4.14)$$

where $C(x)$, $C_s(x)$ and $C_t(x)$ are the given capacity functions and $\nabla \cdot q$ calculates the total spatial flow nearby x , in comparison with the summation operator stated in 4.9. Similar to the discrete max-flow problem in 4.10, the continuous max-flow model can be described as:

$$\max_{q_s, q_t, q} \int_{\Omega} q(x) dx \quad (4.15)$$

by taking into account the constraints of 4.11, 4.12, 4.13 and 4.14. The model in 4.15 is described as the “primal model” in [2] and all flow functions q_s , q_t and q are known as “primal variables”. Through the introduction of a multiplier known as the “dual variable” and denoted by μ to the flow conservation of 4.14, the continuous maximal flow model 4.15 can be written as its corresponding “primal-dual” model:

$$\max_{q_s, q_t, q} \min_{\mu} \int_{\Omega} q_s dx + \int_{\Omega} \mu (\nabla \cdot q - q_s + q_t) dx \quad (4.16)$$

such that, $q_s(x) \leq C_s(x)$, $q_t(x) \leq C_t(x)$ and $|q(x)| \leq C(x)$. The model in 4.16 can also be described as the lagrangian function of 4.15. It is possible to define its respective augmented lagrangian function, F_c , as:

$$F_c(q_s, q_t, q, \mu) := \int_{\Omega} q_s dx + \int_{\Omega} \mu (\nabla \cdot q - q_s + q_t) dx - \frac{c}{2} \|\nabla \cdot q - q_s + q_t\|^2 \quad (4.17)$$

where $c > 0$. From here, an algorithm is developed for the continuous max-flow problem based on the augmented lagrangian method [3] where μ is updated as the multiplier at each iteration.

The segmentation of an image in an unsupervised manner begins by taking two grey values, g_1 and g_2 , which are selected in order to generate the following data terms:

$$C_s(x) = D(g(x) - g_1(x)) \quad (4.18)$$

$$C_t(x) = D(g(x) - g_2(x)) \quad (4.19)$$

where $D(\cdot)$ is some penalty function.

Next, initial values are assigned to q_s^1 , q_t^1 , q^1 and μ^1 and let the current iteration value, $k = 1$. For every k -th iteration the following steps are taken as described in Algorithm 4.1 until convergence.

Algorithm 4.1 Multiplier-Based Maximal-Flow Algorithm

- Update and optimise q :

$$\begin{aligned} q^{k+1} &:= \arg \max_{\|q\|_\infty \leq C(x)} F_c(q_s^k, q_t^k, q, \mu^k) \\ &= \arg \max_{\|q\|_\infty \leq C(x)} -\frac{c}{2} \|\nabla \cdot q(x) - H^k\|^2 \end{aligned}$$

resulting in a projection approach, where H^k is a fixed variable.

- Update and optimise q_s :

$$\begin{aligned} q_s^{k+1} &:= \arg \max_{q_s(x) \leq C_s(x)} F_c(q_s, q_t^k, q^{k+1}, \mu^k) \\ &:= \arg \max_{q_s(x) \leq C_s(x)} \int_{\Omega} q_s dx - \frac{c}{2} \|q_s - I^k\|^2 \end{aligned}$$

where I^k is a fixed variable and optimising q_s can be computed at each point, $x \in \Omega$.

- Update and optimise q_t :

$$\begin{aligned} q_t^{k+1} &:= \arg \max_{q_t(x) \leq C_t(x)} F_c(q_s^{k+1}, q_t, q^{k+1}, \mu^k) \\ &:= \arg \max_{q_t(x) \leq C_t(x)} -\frac{c}{2} \|q_t - J^k\|^2 \end{aligned}$$

where J^k is a fixed variable and optimising q_t can be solved by:

$$q_t(x) = \min(J^k(x), C_t(x))$$

- Update μ :

$$\mu^{k+1} = \mu^k - c(\nabla \cdot q^{k+1} - q_s^{k+1} + q_t^{k+1})$$

- Let $k = k + 1$ return to the $(k + 1)$ -th iteration until convergence.
-

4.3.1 Edge Detection and Boundary Matching

The segmented image volume undergoes a transformation via structured forest learning [4] where the boundaries or edges of pancreatic issue and surroundings are detected. The edges of segments in each slice are measured against the boundaries of equally sized pancreas segments provided in the training dataset. The measures of similarity between these edges are performed via modified Hausdorff distance (MHD) [5] and structural similarity (SSIM) index [6].

An MHD measure provides a relatively local comparison between two pancreatic boundary points in comparison to SSIM, which tends to capture a much more global similarity between two image patches. For example, an MHD captures small circular patterns or strokes of pancreatic contours, whereas SSIM provides an overall morphology “appearance” similarity between contours being compared. Hence, combining these two metrics provides the advantages of both global and local measures of similarity.

Whenever the error between a region in the training data and its corresponding region in the segmentation slice falls below empirically identified threshold values of 0.15 (MHD) and 0.29 (SSIM), a boundary match is assigned to a compilation of pancreas contour similarities, otherwise the contour is discarded. Figure 4.2(e) depicts the boundaries of different tissue in a segmented slice after max-flow and min-cuts segmentation. Notice the variation in contour intensity against the background: the greater outline of the pancreas has been heavily detected whereas the edges inside are less enhanced, thus preventing unnecessary segment division in later processing.

Structured Forest Edge Detection Algorithm

Edge detection via structured forests (SE) employs structured learning to address problems of edge patterns in the pancreas region that display localised characteristics resembling straight lines and curves. In any given 2D image, the edges in a local image patch are highly co-dependent. A set of learning approaches called structured learning are employed, in which a random decision forest approach is utilised to extract the structured information in edge patches. Consider a decision tree, $d_t(x)$ that organises a sample $x \in X$ by iteratively branching left or right down the tree until reaching a leaf node. If a node is denoted as i , then every i in the tree is associated with a binary *split function*, parameterised by θ_i and denoted by:

$$h(x, \theta_i) \in \{0, 1\} \quad (4.20)$$

Should $h(x, \theta_i) = 0$, then node, i pushes x to the left, otherwise to the right, with the operation ending at a leaf node. The output of the tree on an input x is the prediction

stored at the leaf reached by x , which could represent a target label, $y \in Y$. A decision forest is an ensemble of T independent trees d_t . For a given sample x , the predictions $d_t(x)$ from the set of trees are merged using an ensemble model into a single output. It should be noted that random, unusual information may be stored at the leaves of a decision tree. The leaf node reached by the tree is dependent only upon the input x , and although predictions of multiple trees must be combined using an ensemble model, any type of output y can be stored at each leaf. This, consequently, permits the use of complex output spaces, Y , including structured outputs. Next, the training of random decision forests with structured Y is discussed. Each tree is trained independently and recursively. For a given node i and training set $S_i \subset X \times Y$, the objective is to find parameters θ_i of the split function $h(x, \theta_i)$ that yield a split that groups similar segments. This, consequently, requires an *information gain* defined as:

$$I_i = I(S_i, S_i^L, S_i^R) \quad (4.21)$$

where $S_i^L = \{(x, y) \in S_i \mid h(x, \theta_i) = 0\}$ and $S_i^R = S_i \setminus S_i^L$. Choosing to split parameters θ_i maximises the information gain and then training proceeds iteratively on the left node with data, S_i^L , and similarly for the right node with data, S_i^R . The training terminates when the value of the depth reaches maximum or when the information gain drops below a fixed threshold.

The main objective of edge detection is to predict local segmentation masks for a local image patch. The SE approach to learning random decision trees uses structured labels to determine the splitting function at each branch in the tree. Afterwards, the structured labels are mapped to a discrete space, from which information gain is evaluated. Every forest makes a prediction of patches relating to edge pixel labels that are combined across the image to yield the final edge map. The purpose of structured learning is to tackle the problem of optimising or “learning” a mapping where the input or output space may be indiscriminately complex, such as the interdependence of edges in an image.

In testing a trained SE model, a 2D image is processed and a label is assigned to each pixel with a binary variable, denoting whether the pixel contains an edge or otherwise. There is the assumption that a set of segmented training images is provided, in which the boundaries between the segments relate to contours. The method described in this chapter integrates a pretrained SE model based on the BSDS500 natural image dataset [208], and successfully produces detailed edge detection as shown in Figure 4.2.

Boundary Matching Algorithms

After a 2D image (slice) from the current 3D segmentation has been converted to edges, a small image region from that slice is measured against equally sized image regions containing closed pancreatic contours from multiple slices in the training image volumes. Firstly, the modified Hausdorff distance (MHD) is computed. Consider two 2D images of the same spatial size, \mathbf{A}_H and \mathbf{B}_H , where the former refers to a local slice region in the 3D segmentation outcome and the latter refers to a corresponding slice region of an image volume in the training dataset. Both images can be represented as point sets, $\mathbf{A}_H = \{a_1, \dots, a_{N_a}\}$ and $\mathbf{B}_H = \{b_1, \dots, b_{N_b}\}$, respectively, where $\mathbf{A}_H, \mathbf{B}_H \subset \mathbb{R}^n$ such that $|\mathbf{A}_H|, |\mathbf{B}_H| < \infty$. From here, the distance between a point a and set of points, \mathbf{B}_H , is defined as:

$$d(a, \mathbf{B}_H) = \min_{b \in \mathbf{B}_H} \|a - b\| \quad (4.22)$$

and similarly, the distance between a point b and set of points, \mathbf{A}_H , is defined as:

$$d(b, \mathbf{A}_H) = \min_{a \in \mathbf{A}_H} \|b - a\| \quad (4.23)$$

The MHD, D_{MH} , can be described as:

$$D_{MH}(\mathbf{A}_H, \mathbf{B}_H) = \max \left\{ \frac{1}{|\mathbf{A}_H|} \sum_{a \in \mathbf{A}_H} d(a, \mathbf{B}_H), \frac{1}{|\mathbf{B}_H|} \sum_{b \in \mathbf{B}_H} d(b, \mathbf{A}_H) \right\} \quad (4.24)$$

Empirically investigated, it is noted that when $D_{MH}(\mathbf{A}_H, \mathbf{B}_H) = 0$, this often but not exclusively implies that $\mathbf{A}_H = \mathbf{B}_H$; nevertheless, this metric has more discriminatory power than the generalised Hausdorff distances [8].

Secondly, for the structural similarity (SSIM) index measure of similarity, consider two separate and distinct edge-based 2D images, \mathbf{A}_S and \mathbf{B}_S , where the former refers to a local slice region in the 3D segmentation outcome and the latter refers to a corresponding slice region of an image volume in the training dataset. The SSIM index is defined:

$$\text{SSIM}(\mathbf{A}_S, \mathbf{B}_S) = [v_l(\mathbf{A}_S, \mathbf{B}_S)]^\alpha [v_c(\mathbf{A}_S, \mathbf{B}_S)]^\beta [v_s(\mathbf{A}_S, \mathbf{B}_S)]^\gamma \quad (4.25)$$

where the metric lies in the range $-1 \leq \text{SSIM}(\mathbf{A}_S, \mathbf{B}_S) \leq 1$ and where $\text{SSIM}(\mathbf{A}_S, \mathbf{B}_S) = \text{SSIM}(\mathbf{B}_S, \mathbf{A}_S)$. It should be noted that in most cases, a score is calculated in the interval $[0, 1]$, where values closer to 0 represent lower levels of image similarity and values closer to 1 indicate higher levels of image similarity. The terms $\alpha > 0$, $\beta > 0$ and $\gamma > 0$

influence the relative significance of each of the three terms of the SSIM index. The luminance, v_l , contrast, v_c , and structural component, v_s , of the index are defined as:

$$v_l(\mathbf{A}_S, \mathbf{B}_S) = \frac{2\lambda_x\lambda_y + K_1}{\lambda_x^2 + \lambda_y^2 + K_1} \quad (4.26)$$

$$v_c(\mathbf{A}_S, \mathbf{B}_S) = \frac{2\mu_x\mu_y + K_2}{\mu_x^2 + \mu_y^2 + K_2} \quad (4.27)$$

$$v_s(\mathbf{A}_S, \mathbf{B}_S) = \frac{\mu_{xy} + K_3}{\mu_x\mu_y + K_3} \quad (4.28)$$

where λ_x and λ_y represent the mean values of the original images, \mathbf{A}_S and \mathbf{B}_S , respectively; μ_x and μ_y represent the standard deviations, respectively; μ_x^2 and μ_y^2 denote the variances, respectively; and μ_{xy} is the co-variance of the two images. In order to cope with the situations in which the denominators are close to zero, the constants K_1 , K_2 and K_3 are introduced, in which $K_1 = (Z_1G_L)^2$, $K_2 = (Z_2G_L)^2$ and $K_3 = \frac{K_2}{2}$. Here, an 8-bit image has a dynamic range of pixel values, $G_L = 255$ and $Z_1 = 0.01$ and $Z_2 = 0.03$ [7].

4.4 Refinement

Once a rough segmentation has been extracted, a stage of post-processing eliminates surrounding contours identified as “non-pancreas”. Figure 4.2(f) displays the final segmentation outcome for a single slice in a test image volume following three levels of refinement. Each level progressively targets surrounding tissue that are located in greater proximity to the pancreas. The first level performs a shallow-based removal of distinct non-pancreatic contours such as whole or remaining fractions of the aorta, portal vein, duodenum, stomach, and the ascending and descending colon. The second level tackles non-pancreatic tissue that does not lie within the radiologically described embodiment of a pancreas shape [9, 10]. Last, but certainly not least, the third level of refinement analyses the positioning of distinct contours that are very close to but are not part of the pancreas’ head and body, such as the inferior vena cava, common hepatic artery, splenic artery and vein, superior mesenteric, a fraction of the stomach; and close to but not part of the pancreas’ tail including a fraction of the left kidney and spleen. Depending on the slice position in the segmented image volume, such tissue can be separated by just a few pixels (in the range of [1, 5]) from pancreatic tissue and thus requires further analysis to increase the likelihood of correct tissue classification. In order to emphasise the importance of this post-processing stage, Figure 4.3 highlights the major pancreas region in three slices (from three different image volumes) that progress through three consecutive levels of refinement. Notice that close-range non-pancreatic contours are effectively removed in the third level.

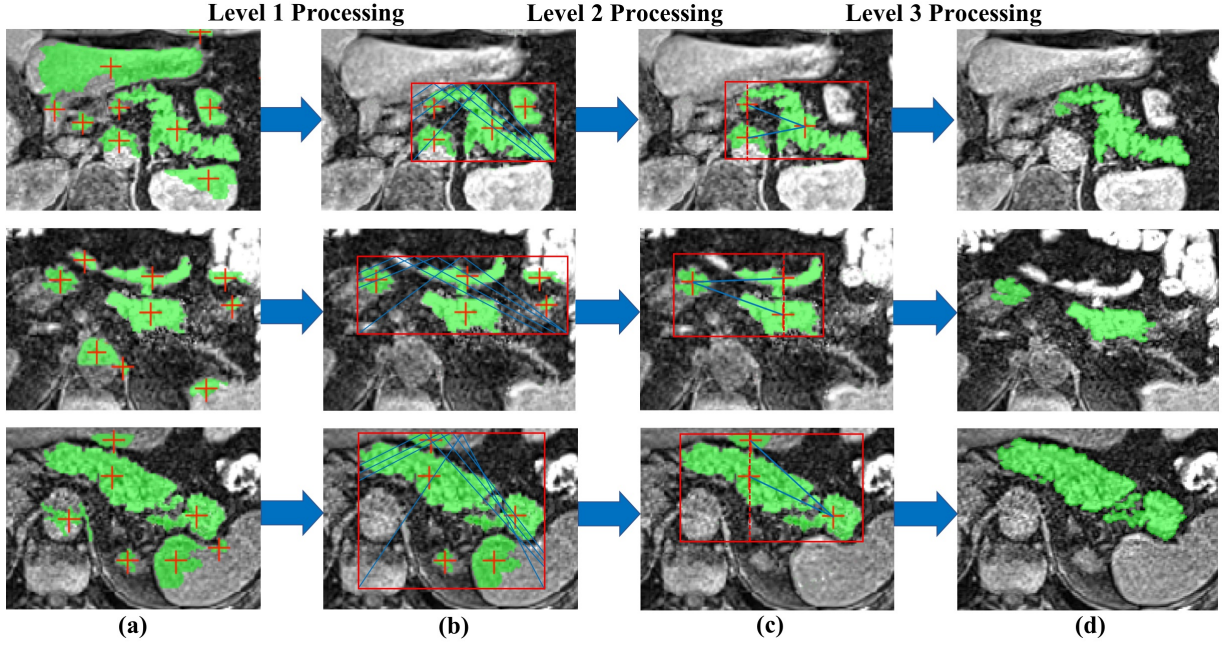


Figure 4.3: Visualisation of three levels for fine pancreas extraction in three different slices from three different image volumes. Column (a) displays the original slice after initial segmentation. Column(b) displays the resultant slice after first level of refinement, now contained in a bounding box against a “trail map”. Column (c) displays the resultant slice after second level of refinement, now contained in a bounding box that identifies gradient between combinational distinct contours. Column (d) displays final, resultant slice after third level of refinement.

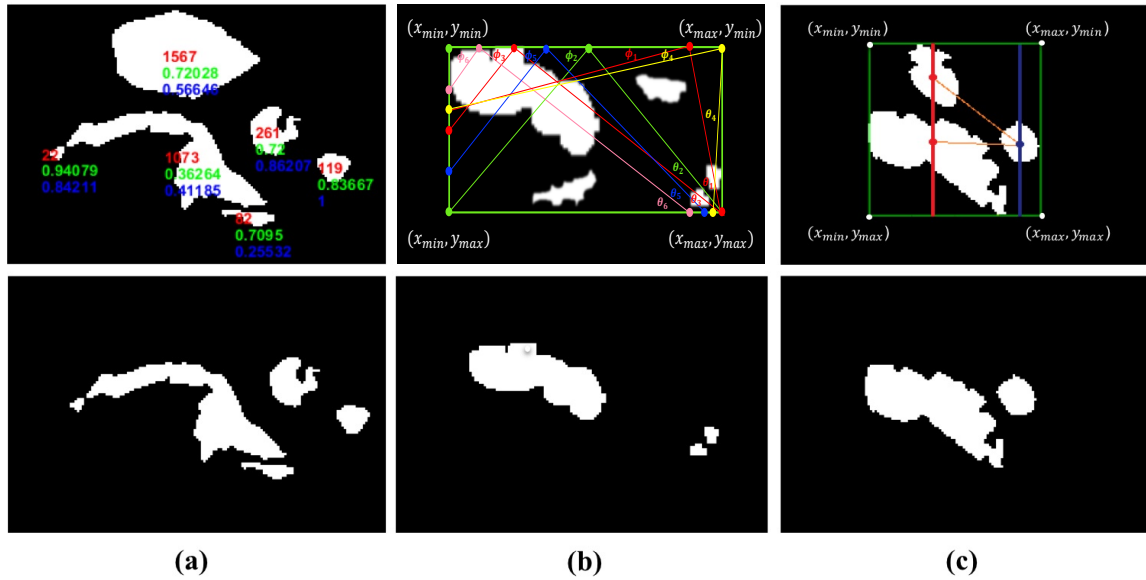


Figure 4.4: Visualisation of three refinement levels in three different slices (in three different image volumes). Top row and bottom row display distinct contours before and after processing. Column (a) displays first level of morphological operations on distinct contours, highlighting measurements of area, spatial aspect ratio and triangularity. Column (b) displays second level of positioning contours on “trail map” within bounding box containing all contours. Column (c) displays combinational connectivity between centre landmark points and respective gradients between all distinct contours.

4.4.1 Level 1: Morphological Operations on Distinct Contours

Analysing a given set of annotated training image volumes, the following is deduced: the mean of ranges (where the pancreas is visible throughout successive slices in a volume), with careful consideration to heavy outliers; the mean slice number where the pancreas features become visible for the first time, s_{start} , reach maximum area, s_{max} , and the last slice after which pancreas features are not visible any longer, s_{end} . By considering the total number of slices in each image volume, t_s , a discrete set of four slice ranges are established: $r_1 : [1, s_{start}-1]$, $r_2 : [s_{start}, s_{max}-1]$, $r_3 : [s_{max}, s_{end}-1]$ and $r_4 : [s_{end}, t_s]$, where r_1, r_2, r_3 and $r_4 \in \mathbb{Z}$. A k -medoids cluster approach is employed in order to generate $N_{r_1}, N_{r_2}, N_{r_3}$ and N_{r_4} number of groups of constraints for every slice range. Every group of constraints includes a measure of area, triangularity and ratio of spatial dimensions. Let $N_{r_i}(\alpha)$, $N_{r_i}(\beta)$ and $N_{r_i}(\gamma)$ represent individual constraints of area, triangularity and ratio of spatial dimensions. Thus, for an observed segmentation contour that corresponds to N_{r_i} , the following operation can be defined in 4.29 as:

$$p_c = \begin{cases} 1 & \text{if } 0 \leq E_{r_i}^\alpha, E_{r_i}^\beta, E_{r_i}^\gamma \leq t_m \\ 0 & \text{otherwise} \end{cases} \quad (4.29)$$

where a value of 1 corresponds to “pancreas” and a value of 0 corresponds to “non-pancreas”. $E_{r_i}^\alpha$, $E_{r_i}^\beta$ and $E_{r_i}^\gamma$ represent the error between $N_{r_i}(\alpha)$, $N_{r_i}(\beta)$ and $N_{r_i}(\gamma)$ and an observed segmentation contour’s similar measures, respectively. If the error is greater than t_m , then this contour is removed from the segmentation result.

Another morphological operation involves computation of mean curvature [11] of distinct contours in each segmentation slice. Suppose that C_p represents a contour. Let the angle between the tangents to C_p at points q_1 and q_2 be $\theta(q_1, q_2)$, and let the length of the segment of the curve between q_1 and q_2 be $s(q_1, q_2)$. The curvature, κ_p , of the curve C_p to q_1 is defined in Equation 4.30 as:

$$\kappa_p = \lim_{q_2 \rightarrow q_1} \frac{\theta(q_1, q_2)}{s(q_1, q_2)} \quad (4.30)$$

If the curvature value, κ_p , of an observed contour falls below a threshold, t_c , which is based on the curvature analysis of unique contours in the training dataset, then it is discarded from the resultant segmentation. The top image in Figure 4.4(a) illustrates an example where several distinct contours in a slice are labelled with area, ratio and triangularity; the bottom image displays the resultant slice after removal of contours deemed as “non-pancreas”.

4.4.2 Level 2: Localisation and Positioning of Contours

The slice-by-slice inspection of pancreatic regions in the training dataset reveals that whole or distinct pancreatic contours are embodied in a shape resembling a horseshoe, an inverted-V, transverse, sigmoidal [9] but more commonly, oblique or L-shaped [10]. Therefore, consider a localisation bounding box to contain all the contours in each segmentation slice, as described in 4.31:

$$F_s = [x_{min}, y_{min}, (x_{max} - x_{min}), (y_{max} - y_{min})] \quad (4.31)$$

where (x_{min}, y_{min}) represents the top left-vertex of the bounding box, $(x_{max} - x_{min})$ is the width and $(y_{max} - y_{min})$ is the height that follows from this point of reference. From here, it is possible to generate an L-shaped template that behaves like a “trail-map” for identifying contours deemed as “pancreas” or otherwise. This trail-map can be viewed as a collection of neighbouring paths that begin from a set of points, $B_{XY} = \{(x_1, y_{max}), \dots, (x_n, y_{max})\}$, on the bottom horizontal of the bounding box and rise by corresponding angles, $\theta = \{\theta_1, \dots, \theta_n\}$, to respective points on the top horizontal of the bounding box. From here, the trail descends by angles, $\phi = \{\phi_1, \dots, \phi_n\}$, to respective terminating points (on the bounding box). It is noted that $B_X = \{x_1, \dots, x_n\}$ are values that refer to a set of n distances measured from the bottom right-hand vertex, i.e. (x_{max}, y_{max}) , hence, $B_X \in \mathbb{R} \mid x_{min} \leq B_X \leq x_{max}$. Values of θ and ϕ are co-dependent on the width and height of the bounding box.

Taking into consideration the variation in pancreatic contours, a further trail resembling an inverted-V is added: the trail begins at a point that lies mid-way on the right vertical of the bounding box, $(x_{max}, \frac{1}{2}(y_{max} + y_{min}))$, meets the point mid-way on the top horizontal line, $(\frac{1}{2}(x_{max} + x_{min}), y_{min})$ and then descends to a point that lies mid-way on the left vertical line of the bounding box, $(x_{min}, \frac{1}{2}(y_{max} + y_{min}))$. The top image in Figure 4.4(b) illustrates an example where several oblique trails run across distinct segments. If an observed contour from a segmentation slice does not lie on the trail-map, it is removed from the image volume as displayed in the bottom image of Figure 4.4(b), otherwise it is retained to progress to the next post-processing level.

4.4.3 Level 3: Centre Landmarks of Distinct Contours

A final level of refinement examines the gradient between pairs of contours. For example, in any given slice in the training dataset, it is unusual, if not uncommon to observe two pancreatic contours where the gradient between their landmark centre points is infinity. Therefore, if the centre points of two pancreatic contours in a segmentation slice lie on the same imaginary vertical line, then it is safe to suggest that one of two is likely “non-pancreas” and opt for removal. The top image in Figure 4.4(c) illustrates an example

where a pair of contours with gradient infinity have been identified; the bottom image displays the resultant slice after removal of the contour deemed “non-pancreas”.

In order to identify which contour to eliminate from the resultant segmentation, a number of computations are performed to generate multiple groups of “segmentation characteristics”. These groups are analysed against multiple groups of similar measurements in the training data, otherwise known as “ground-truth characteristics”.

Generate groups of “ground-truth characteristics”

Firstly, for every image volume in a training dataset, the slices that contain two, three or four distinct pancreas contours are analysed. For each slice, it is possible to generate its respective “ground-truth characteristics” that contain: the areas of individual contours; and the gradient, Euclidean distance and ratio-of-areas between pairs of distinct contours, all computed in a combinational manner.

For example, as highlighted in Figure 4.5(a), if an image slice has three distinct contours with areas, A , B and C , then three combination gradients, g_{AB} , g_{AC} and g_{BC} are computed, as well as corresponding Euclidean distances, d_{AB} , d_{AC} and d_{BC} . Furthermore, the ratios of the former area to the latter, i.e., A/B , A/C , and B/C , are important descriptors, providing an indication of overall contour shape as opposed to area alone.

Four groups, R_1 to R_4 , describing “ground-truth characteristics” of different combinations of contours, are generated. Each group is confined to its respective slice range, r_1 to r_4 . As highlighted in Figure 4.5(b), R_1 describes a set containing groups of “ground-truth characteristics” from a combination of two contours. R_2 describes two sets, one set has groups of “ground-truth characteristics” from a combination of two contours, and another set has groups of “ground-truth characteristics” from a combination of three contours. R_3 describes three sets, each set contains groups of “ground-truth characteristics” from a combination of two, three and four distinct contours separately.

Each set of groups of “ground-truth characteristics” can be represented by f_p^i where i refers to the i -th slice range group, R_i , and p refers to the number of contours. Figure 4.5(b) provides a visual summary of all sets and defined below is an example of f_3^2 :

$$\begin{aligned}
f_3^2 = & \{(A, B, C, g_{AB}, g_{BC}, g_{AC}, d_{AB}, d_{AC}, d_{BC}, A/B, A/C, B/C)_1 \\
& (A, B, C, g_{AB}, g_{BC}, g_{AC}, d_{AB}, d_{AC}, d_{BC}, A/B, A/C, B/C)_2 \\
& (A, B, C, g_{AB}, g_{BC}, g_{AC}, d_{AB}, d_{AC}, d_{BC}, A/B, A/C, B/C)_3 \\
& \dots \\
& (A, B, C, g_{AB}, g_{BC}, g_{AC}, d_{AB}, d_{AC}, d_{BC}, A/B, A/C, B/C)_n\} \quad (4.32)
\end{aligned}$$

where n is a chosen number of groups in a set that is based on k -medoids clustering of these measurements in the training dataset.

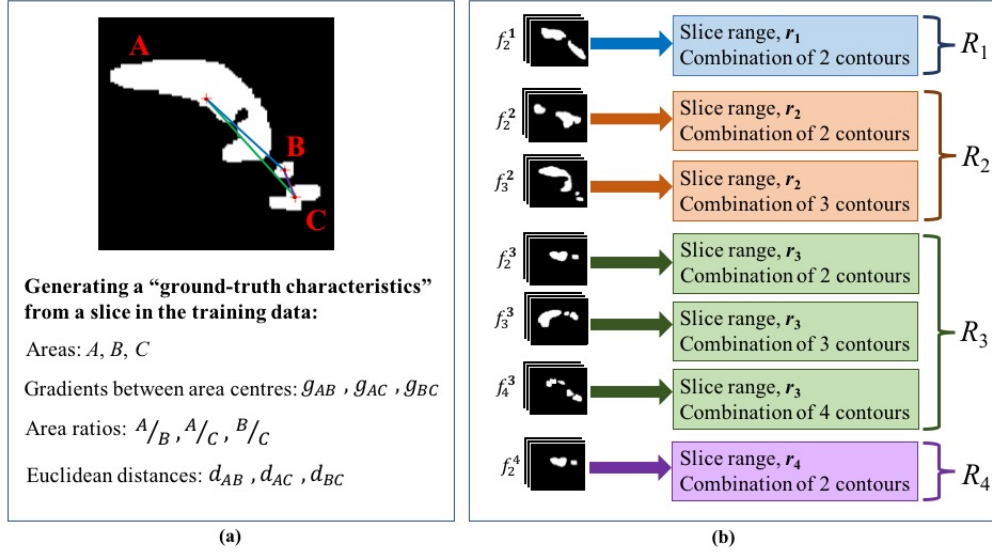


Figure 4.5: Column (a) provides a visual example of a slice from an image volume in a training dataset. Computations include the gradient and distance between combinational area centres and the respective ratios of combinational areas. Column (b) provides a summarised visual representation of “ground-truth characteristics” generated for sets R_1 , R_2 , R_3 and R_4 .

Generate groups of “segmentation characteristics”

It should be noted that the focus lies on analysing two, and only two contours in each segmentation slice whose centres lie on the same imaginary vertical line. This limits computational time, but the general technique can be extended to a greater number of contours. Thus, in each segmentation slice, let M represent the total number of separate contours and let N represent the total number of vertical lines that have two contours, i.e., where the gradient between respective centre points is infinity. Therefore, it is of interest to retain $T = (M - N)$ contours in the slice. Next, the total number of possible combinations of T from M contours is computed, as denoted by C_{TM} and defined in Equation 4.33 as:

$$C_{TM} = \frac{M!}{T!(M - T)!} \quad (4.33)$$

For every $C_{TM}(j)$ -th combination of contours, where $j \in \mathbb{Z} : j \in [1, C_{TM}]$, the gradients of all possible paired contours are calculated, in addition to respective distances and ratios.

For example, given a slice range group R_3 , if $M = 5$ and $T = 3$, then $C_{3,5} = 10$ different combinations of three contours from five contours (where two are presumed invalid). It is important to note that a number of combinations will be irrelevant since the gradient between two centres is infinity, and thus can be discarded (see Appendix E for further information). Hence, for every remaining (relevant) three contour combination, e.g. let

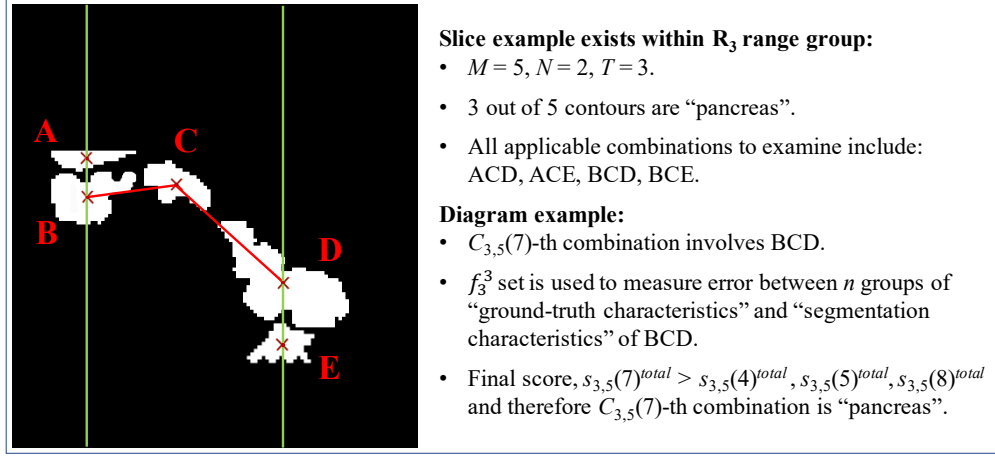


Figure 4.6: Visual example of three contour combination, BCD , which represents the 7th in 10 different combinations from five distinct contours. Three out of five contours are presumed valid, i.e. “pancreas” and the other two otherwise.

$C_{3,5}(7)$ involve areas, B, C, D , a “segmentation characteristics” group is generated, consisting of three gradients, areas, Euclidean distances and ratios-of-areas. This group is simultaneously evaluated against a set of “ground-truth characteristics” from the same slice range group and T contours of interest. Figure 4.6 provides a visual representation of this example, where set f_3^3 would be utilised (see Appendix F for further detail regarding this example).

A method of elimination follows: if the mean error between the $C_{TM}(j)$ -th “segmentation characteristics” group and a “ground-truth characteristics” group is equal or falls below threshold, t_p , then this combination is assigned a point score of 1, otherwise 0. Once all measures of error have been exhausted, the combination with the highest score is retained as “pancreas”, and the others discarded. Equation 4.34 mathematically describes the process of assigning a score to the j -th combination, $s_{TM}(j)^k$, where $k \in \mathbb{Z} : k \in [1, n]$.

$$s_{TM}(j)^k = \begin{cases} 1 & \text{if } 0 \leq E_{f_p^i}^k \leq t_p \\ 0 & \text{otherwise} \end{cases} \quad (4.34)$$

where $E_{f_p^i}^k$ is the mean error between the $C_{TM}(j)$ -th combination “segmentation characteristics” and $(f_p^i)_k$ which is the k -th “ground-truth characteristics”. The final score, $s_{TM}(j)^{total}$ can be described in Equation 4.35 as:

$$s_{TM}(j)^{total} = \sum_{k=1}^n s_{TM}(j)^k \quad (4.35)$$

In the case where two or more combinations have attained the same score, the combination that encapsulates these two contours scoring the least mean error is retained.

4.5 Conclusion

In this chapter, an automatic approach is presented for pancreas segmentation in 3D abdominal computed tomography (CT) and Magnetic resonance imaging (MRI) scans. The proposed method is based on a hierarchical pooling of information by classifying extracted image patches, superpixels and intensity distributions as pancreatic tissue or otherwise. A sequential process firstly improves the difference between pancreatic and background classes using optimised parameters from a learned intensity model. With focus on differentiating the nearby splenic vein, superior mesenteric artery and surrounding tissue that are often in close proximity to the pancreas, the proposed digital contrast intensity model improves the distributional characterisation of intensities between pancreatic and surrounding tissue boundaries. Next, the major pancreas region and surrounding tissue are extracted using a hybrid segmentation approach: the major pancreas region is identified using a trained random forest that makes probability-wise predictions of superpixels in each slice of the image volume. This technique, followed by 3D max-flow and min-cuts segmentation and structured forest edge detection, produces very detailed contouring of the pancreas and heavy boundary-preservation. This voxel-based algorithm addresses the intensity consistency problem that is often the case when segmenting image volumes on a slice-by-slice basis. Furthermore, a novel post-processing stage optimises tissue classification using morphological, anatomical and radiological knowledge about connectivity between pancreatic contours in an image volume.

Chapter 5

Method 2: Integrating Deep Learning

5.1 Introduction

The success of deep learning or convolutional neural networks (CNNs) for object classification has motivated researchers to take advantage of their feature learning capabilities for pixel and voxel-wise prediction problems such as organ segmentation, and in particular organs with high structural variability. This chapter presents a novel approach (Method 2) for automatic pancreas segmentation in Magnetic resonance imaging (MRI) and Computer tomography (CT) scans. Building on a significant part of Method 1 in Chapter 4, this alternative approach exploits the advantages of deep learning to deliver pixel-wise predictions of scanned abdominal tissue as “pancreas” or “non-pancreas”. This section provides an overview of the second proposed approach (Method 2), in which there are two main phases: training and testing. The training phase firstly identifies a major pancreas, after which a deep learning network is trained to recognise pancreatic features and boundaries. The testing phase integrates the fully trained deep learning model, that, when coupled with 3D segmentation and post-processing, produces the final pancreas segmented volume. Section 5.2 addresses the motivation for firstly isolating a major pancreas region. This section also discusses the primary segmentation approach that incorporates a deep learning model for enhanced pixel-wise classification, and presents a novel loss function for improved network training. Section 5.3 discusses the advantage of performing 3D segmentation on the predicted pancreas region in an (unseen) test image volume. The motivation, here, is to separate pancreatic contours from surrounding tissue that is close to but does not touch the pancreas. Section 5.4 discusses the shift from a 2D-based deep learning architecture to a 3D-based network (known as 3D-Method 2), and the likely implications on computational efficiency and segmentation performance. Section 5.5 concludes this chapter including a summary of Method 2 and 3D-Method 2, and key outcomes that will be discussed in the proceeding chapter. In fact, Chapter 6, Section 6.4.3 and Chapter 7, Section 7.2 will compare

and discuss the results obtained in Method 1 and Method 2 and highlight that, indeed, Method 2 surpasses segmentation accuracy scores in DSC and is statistically more robust.

The methodology of the proposed approach, as illustrated in Figure 5.1, consists of a training phase and a testing phase, each one of which is discussed in proceeding sections. The training phase consists of two main parts. The first part identifies a major pancreas region in all training image volumes. Here, a trained random forest performs probability-wise predictions of superpixels in each 2D image (slice) to recognise the major pancreatic area of interest. In the second part, a deep learning model trains for semantic segmentation using the training data containing the major regions of interest, for pancreas feature and boundary recognition. For network training, a novel loss function integrates the modified Hausdorff distance metric and a sinusoidal component. Afterwards, the testing phase consists of three main stages. (1) the fully trained deep learning model predicts the likelihood of every pixel in a test image volume as “pancreas” or “non-pancreas”. (2) an application of digital contrast enhancement follows 3D segmentation on this predicted pancreas volume via continuous max-flow and min-cuts approach. This stage separates possible nearby vessels and tissue that lie close to this organ of interest. (3) non-pancreatic contours (false-positive pancreas contours) are identified and eliminated from the resultant segmented volume via increasing levels of refinement including morphological operations on area, curvature, position and gradient between distinct contours in the segmented volume.

5.2 Training Phase

This section details the training phase, which develops a deep learning model for predicting pixels that represent true-positive pancreas in a given image volume. For every separate dataset (that corresponds to a different MRI sequence¹ or modality²) a separate deep learning model is generated. During project development, the datasets of diverse sequences (i.e. different scanner imaging protocols) and modalities were not mixed. The main motivation for this decision is to maintain a degree of consistency in the structural spatial dimension of the image volumes, and limit the variation of greyscale intensity distribution to each dataset.

¹An MRI sequence corresponds to a scanner imaging protocol, where a particular setting of pulse sequences and field gradients are chosen to emphasise a particular appearance in the resultant image.

²Modality: a type of medical imaging technique that utilises a scanning device to produce images of internal physiological and anatomical structures of a patient. Examples of medical imaging modalities include MRI, CT and ultrasound.

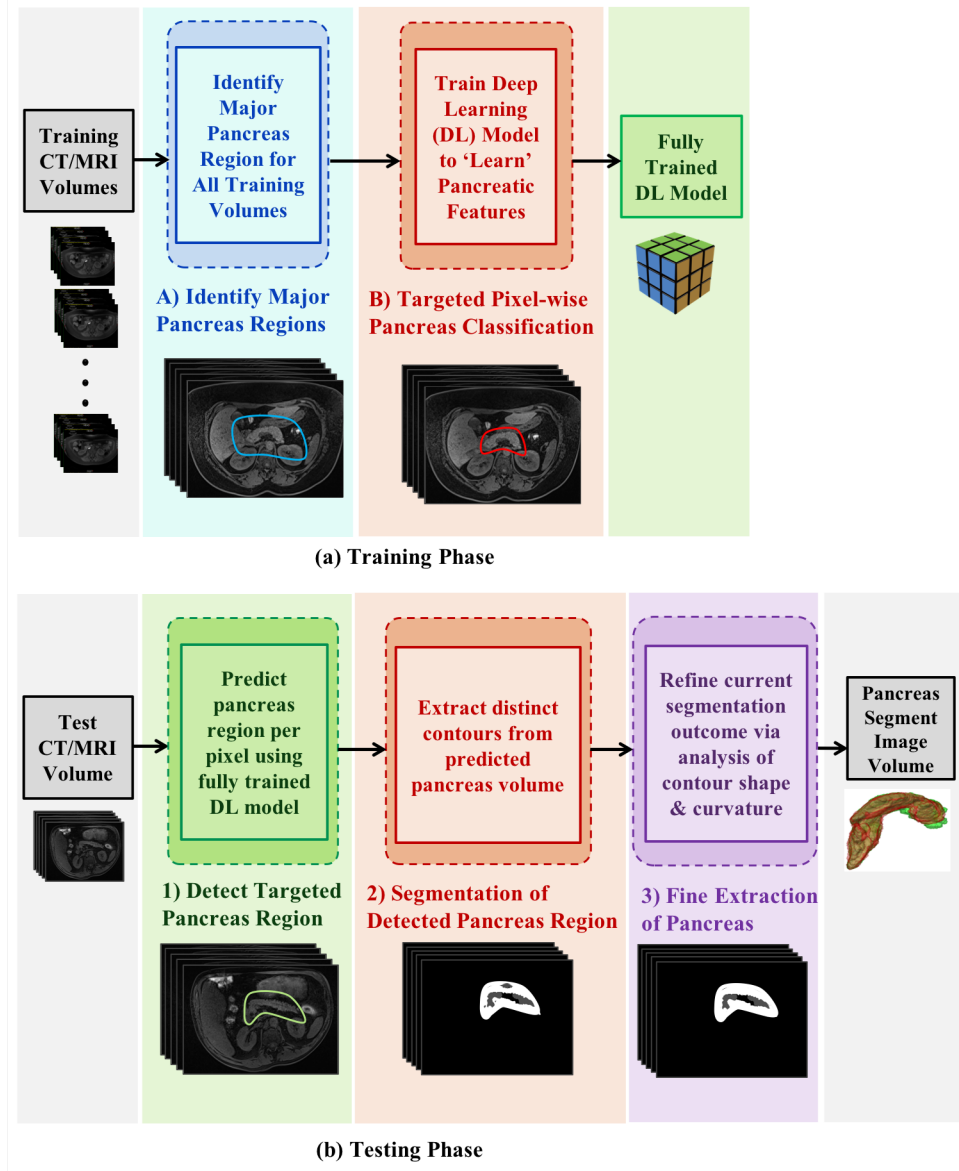


Figure 5.1: Overview of Method 2. Top diagram highlights the training phase consisting of two main parts, (A) train a random forest to identify major pancreas regions in all training image volumes; (B) train a deep learning model to recognise pancreatic features within an image volume, and later perform pixel-wise classification on a test image volume as “pancreas” or “non-pancreas”. The bottom image highlights the testing phase and consists of three main stages, (1) detect targeted pancreas region using the trained deep learning model using pixel-wise classification; (2) perform 3D segmentation of predicted pancreas region to extract distinct contours that may lie near but are not part of the pancreas; (3) perform refinement to identify non-pancreatic contours for removal from final segmented pancreas volume.

5.2.1 Random Forests: Identify Major Pancreas Region

The methodology described in Section 4.2.2 is implemented to firstly identify the greater pancreas region in a test image volume. A random forest is trained on a selection of extracted features in a training dataset of image volumes; afterwards, the trained forest predicts the likelihood of a 32-pixel (region size) superpixel in every slice (2D image) of each image volume as “pancreas” given a probability threshold (0.85). This major region of interest, as a series of slices, feeds into a trained deep learning network (Section 5.2.2) which performs pixel-wise predictions of “pancreas” and “non-pancreas”.

5.2.2 Detect Targeted Pancreas Region

The deep learning architecture of SegNet [100] has proven very useful for semantic segmentation in images that include high-class imbalance (i.e., where the majority of the pixels belong to large classes) and vague boundaries between objects of interest (e.g., road and sidewalk, or, more applicably the pancreas’ head and the duodenum). Thus, SegNet is chosen as a suitable deep learning model to fine-tune and incorporate into the segmentation methodology. This particular CNN has a U-shaped architecture that is associated with an encoder-decoder topology as illustrated in Figure 5.2. The architecture of the encoder network matches the 13 convolutional layers in the VGG16 network [97], whereas the decoder network maps the low-resolution encoder feature maps to full input resolution feature maps for pixel-wise classification. One of the advantages of this deep learning model is the process in which the decoder network upsamples the lower resolution input feature maps. Consequently, this mapping produces features that are useful for defining the minimal rectangle surrounding the pancreas.

Architecture

As highlighted in Figure 5.4, every convolutional layer is achieved using a 3×3 filter (stride of 1 and padding of 1) with initialised training weights [99]. The fully connected layers in VGG16 are discarded to retain higher resolution feature maps at the last encoder output layer. With the encoder network accounting for approximately 14.7 million parameters, each encoder layer has a corresponding decoder layer of 13 layers. The final decoder output is the input to a two-class soft-max classifier to produce probabilities for each pixel, i.e. “pancreas” or “non-pancreas”.

Each encoder in the encoder network performs convolution using a filter (i.e., a set of weights) to generate a set of feature maps, which are then batch normalised [98]. From here, an element-wise rectified-linear non-linearity (ReLU) $\max(0, x)$ is applied [37]. Since the information in an image is a highly non-linear mapping of pixel values in the input, it is important that the mapping from CNN input to its output is also highly non-linear.

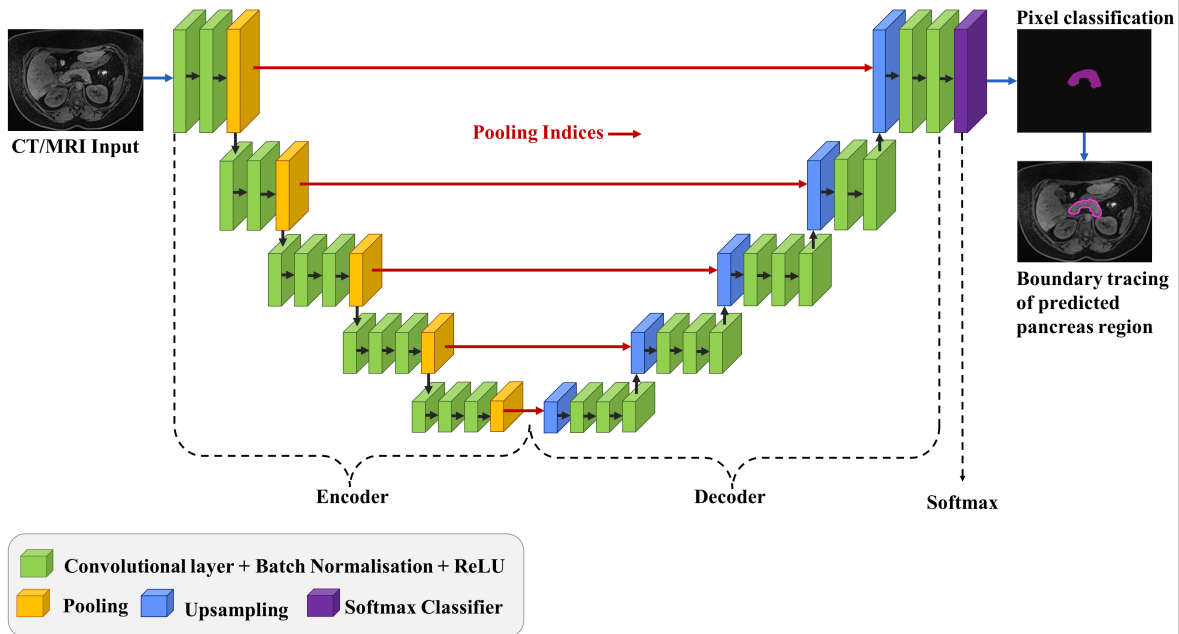


Figure 5.2: Overview of the deep learning model that has been incorporated into Method 2. An encoder downsamples the CT/MRI input through convolution, batch normalisation and ReLU, producing feature maps that represent unique pancreatic features (e.g. texture, boundaries, etc). A decoder upsamples its input using the transferred pooling indices from its corresponding encoder to generate sparse feature maps. From here, convolution is performed with a trainable filter of weights to density the feature map. The resulting decoder output feature maps are fed to a soft-max classifier for 2-channel pixel-wise classification of the input image as “pancreas “ or “non-pancreas”.

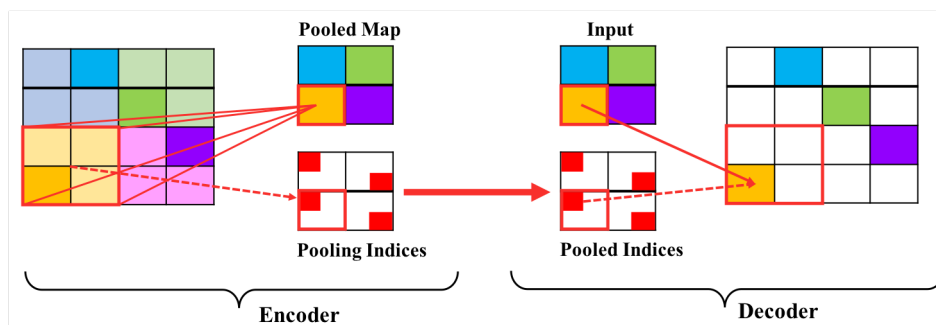


Figure 5.3: A 2×2 max-pooling is applied in the Encoder stage. The maximum value and pooling index inside this 2×2 region is retained and propagated to the next layer. In upsampling, in the Decoder stage, the max-pooled 1×1 feature is placed into to the exact location of its corresponding pooling index.

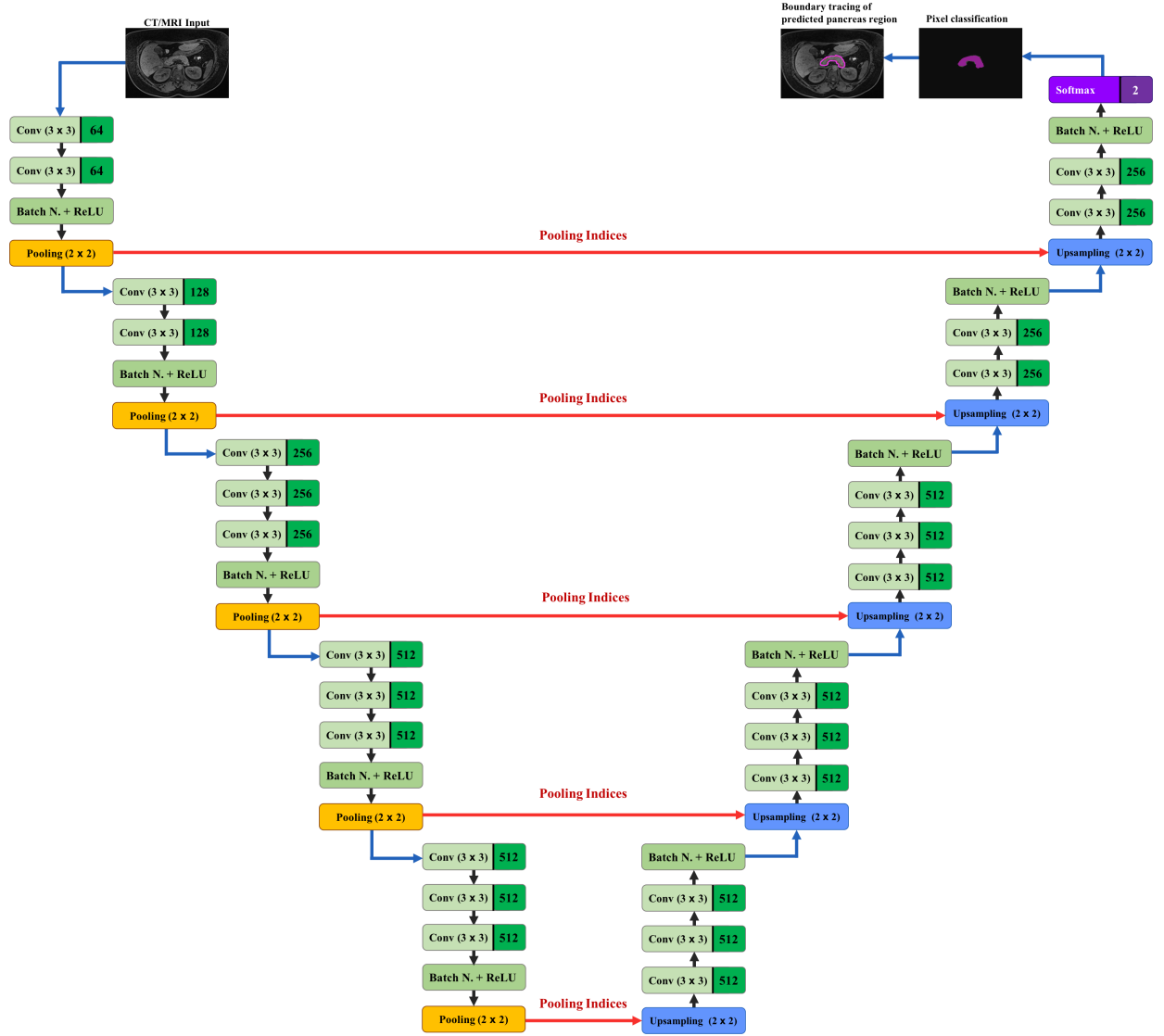


Figure 5.4: Detailed overview of the deep learning architecture. Each round rectangle is a layer, with the number on the right side indicating the number of channels (or feature maps in the case of convolutional layers). Each convolution (*Conv*) layer contains a set of 3×3 filters, stride 1, padding 1. Each max-pooling (*Pooling*) layer uses 2×2 window with stride 2. *Batch N. + ReLU* refers to a batch normalization layer followed by ReLU activation. During max-pooling, the index of the maximum feature in each 2×2 area is saved. For each *Upsampling* layer, the 1×1 feature is placed in the exact location where the corresponding max-pooling index is located. The last layer is a soft-max classifier with 2-channels, one relating to “pancreas” and the other “non-pancreas”. The predicted segmentation relates to the class with maximum probability at each pixel.

Afterwards, a max-pooling layer extracts the maximum value in the output from the previous layer and achieves translation invariance³ over small spatial shifts in the input image. The max-pooling is performed with a 2×2 window and stride 2 (non-overlapping window). While several layers of max-pooling can achieve more translation invariance for accurate classification, there also exists a loss of spatial resolution of the feature maps. Clearly, this loss of boundary detail is not helpful for segmentation where object boundary delineation is essential. Thus, this deep learning model ensures that boundary information is stored in the encoder feature maps prior to any sub-sampling. It involves storing only the max-pooling indices; these are the locations of the maximum feature value in each pooling window, which are “memorised” for each encoder feature map.

The decoder network upsamples its input feature maps using the memorised max-pooling indices from the corresponding encoder feature maps, resulting in thinly dispersed or sparse feature maps as illustrated in Figure 5.3. These feature maps are then convolved with a decoder filter to generate dense feature maps. From here, a batch normalisation layer is applied to each of these feature maps. It is important to note that the decoder relating to the first encoder in the network produces a multi-channel feature map, despite its encoder input having 3 channels (RGB). The high dimensional feature representation at the output of the final decoder becomes the input to a trainable soft-max classifier [101]. The output of the soft-max classifier is a 2-channel image of probabilities with 2 being the number of classes relating to “pancreas” and “non-pancreas”. The predicted segmentation relates to the class with maximum probability at each pixel.

Training the Original (SegNet) Deep Learning Network

As discussed above, the use of max-pooling indices in the decoder stage can perform upsampling of low-resolution feature maps; this has the important advantage of retaining high-frequency details (i.e., improved pancreatic boundary delineation) in the segmented images, and also, to reduce the total number of trainable parameters in the decoders. In training the network, a distinct set of training data depicting the major pancreas region (for each separate dataset) is “fed” into the deep learning model as described above in mini-batches of a defined size. For clarification, the mini-batch is a subset of the training dataset that is used to evaluate the gradient of the loss function and update the weights. By default, the encoder and decoder weights were all initialised using the technique described in [102], and the network was trained using stochastic gradient descent (SGD) with a fixed learning rate of 0.1 and momentum of 0.9. The cross-entropy loss function [68] is used for training and optimising the network, and can be defined as:

³Translation invariance means that an object of interest can still be recognised even when it is translated in some way.

$$C_{loss} = -\frac{1}{n_t} \sum_{i=1}^{n_t} [t_i \log o_i + (1 - t_i) \log(1 - o_i)] \quad (5.1)$$

where n_t is total number of pixels in a training image slice; and o_i and t_i is the corresponding neural network prediction and desired output, respectively.

Next, the deep learning network is updated by:

$$\frac{\partial C_{loss}}{\partial o_i} = -\frac{1}{n_t} \left(\frac{o_i - t_i}{o_i(1-o_i)} \right) \quad (5.2)$$

Improvements using Class Balancing

Whenever there is a large variation in the number of pixels in each class in a training set, which is the case with the pancreas accounting for approximately 1% in a scan, there is a need to weight the loss differently based on the true class (ground-truth label). In order to achieve what is often coined “class balancing”, a method known as “median frequency balancing” [103] is incorporated, where the weight assigned to a class c in the loss function is modified, such that each pixel is weighted by $NP_c^{mfb} = \widetilde{NP}_c / NP_c$, such that NP_c is the number of pixels of class c divided by the total number of pixels in training images where c is present.

Backpropagation

Backpropagation is the process where the network is optimised by updating the convolutional filters (with each filter being a set of weights) and minimising the overall loss between network prediction and target (ground-truth). The learning rate parameter is not kept constant by incorporating a “learning rate drop period” and a “learning rate drop factor”. The learning rate drop period is the number of epochs for dropping the learning rate and the learning rate drop factor is a multiplicative factor that is applied to the learning rate every time a certain number of epochs passes. An epoch corresponds to a full pass of the training data, whereas an iteration number corresponds to a full pass of the mini-batch.

Improvements using Data Augmentation

Computer vision-based classification tasks tend to have insufficient data, and this is particularly true for medical imaging data, especially where access to data is often protected due to privacy concerns. Evidence of effective image classification using data augmentation [108] has supported the development of deep learning techniques that combine expert domain knowledge with pre-trained models. With data augmentation,

the aim is to preserve the main image shape to be recognised. Common data augmentation methods include mirroring on the vertical axis (where an image is flipped horizontally) to produce a new image. Another commonly used technique is random cropping, where fixed-size images are captured from the main training image at different locations. Affine transformations (translation, shearing, rotation) and local warping are also employed to create new data. The term warping was firstly described in [109] by applying both affine transformations and elastic distortions to images in the training data. Another method is colour shifting; for example, the RGB channels can be changed to add more colour variations. In the case of greyscale images, pixel intensity can be reduced or increased within the dark-to-bright range $[0, 255]$. For example, an image could undergo gamma transformation where dark pixels, e.g. $[0, 85]$ are made darker, and brighter pixels are kept the same in intensity. Consequently, colour shifting can make the deep learning algorithm more robust to intensity changes.

Data augmentation can be viewed as a form of “artificial data” generated by applying distortions to the original training dataset. It is used during training to provide more data to the network and help to improve non-linearity and the accuracy of the network. Using such extra data is one way of overcoming overfitting [107]. In this method, random left-right reflection and random X-Y translation of ± 40 pixels (MRI) and ± 80 pixels (CT) are used for data augmentation.

Integrated Hausdorff-Sine Loss Function

A novel loss function is proposed for training the deep learning network through optimising the modified Hausdorff distance (to reduce the boundary matching error) and a sinusoidal functionality that serves to enhance local region and boundary prediction.

In order to define the modified Hausdorff distance-based loss, let T_H and Y_H represent the ground-truth (target) and network boundary (or edge) predictions, respectively; and where $T_H, Y_H \subset \mathbb{R}^n$ such that $|T_H|, |Y_H| < \infty$. Furthermore, t_j and $y_j \in \{0, 1\}$ are indexed pixel values in T_H and Y_H respectively, which can subsequently, be viewed as boundary points. From here, it is possible to define the minimum of the Euclidean (2D) distance between a point t_j and set of points, Y_H as:

$$\begin{aligned} s(t_j, Y_H) &= \min_{y_j \in Y_H} \|t_j - y_j\| \\ &= \min_{y_j \in Y_H} \left(\sqrt{\sum_{k=1}^2 (t_{j_k} - y_{j_k})^2} \right) \end{aligned}$$

and similarly, the minimum distance between a point y_j and set of points, T_H , can be defined as:

$$\begin{aligned}
s(y_j, T_H) &= \min_{t_j \in T_H} \|y_j - t_j\| \\
&= \min_{t_j \in T_H} \left(\sqrt{\sum_{k=1}^2 (y_{j_k} - t_{j_k})^2} \right)
\end{aligned}$$

The loss, L_{mhd} , can be described as:

$$L_{mhd} = \max \left\{ \frac{1}{|Y_H|} \sum_{y_j \in Y_H} s(y_j, T_H), \frac{1}{|T_H|} \sum_{t_j \in T_H} s(t_j, Y_H) \right\} \quad (5.3)$$

Thus, computing the gradient (see Appendix G for breakdown of derivation) yields:

$$\frac{\partial L_{mhd}}{\partial Y_H} = \begin{cases} \frac{\partial}{\partial Y_H} \left(\frac{1}{|Y_H|} \sum_{y_j \in Y_H} s(y_j, T_H) \right) & \text{if } \frac{1}{|Y_H|} \sum_{y_j \in Y_H} s(y_j, T_H) > \frac{1}{|T_H|} \sum_{t_j \in T_H} s(t_j, Y_H) \\ \frac{\partial}{\partial Y_H} \left(\frac{1}{|T_H|} \sum_{t_j \in T_H} s(t_j, Y_H) \right) & \text{if } \frac{1}{|Y_H|} \sum_{y_j \in Y_H} s(y_j, T_H) < \frac{1}{|T_H|} \sum_{t_j \in T_H} s(t_j, Y_H) \\ \text{undefined} & \text{if } \frac{1}{|Y_H|} \sum_{y_j \in Y_H} s(y_j, T_H) = \frac{1}{|T_H|} \sum_{t_j \in T_H} s(t_j, Y_H) \end{cases} \quad (5.4)$$

where,

$$\frac{\partial}{\partial Y_H} \left(\frac{1}{|Y_H|} \sum_{y_j \in Y_H} s(y_j, T_H) \right) = \frac{1}{|Y_H|} \sum_{y_j \in Y_H} \min_{t_j \in T_H} \left(\frac{\sum_{k=1}^2 (y_j - t_j)_k}{\|y_j - t_j\|} \right) \quad (5.5)$$

and

$$\frac{\partial}{\partial Y_H} \left(\frac{1}{|T_H|} \sum_{t_j \in T_H} s(t_j, Y_H) \right) = \frac{1}{|T_H|} \sum_{t_j \in T_H} \min_{y_j \in Y_H} \left(\frac{\sum_{k=1}^2 (y_j - t_j)_k}{\|t_j - y_j\|} \right) \quad (5.6)$$

Building on the conventional cross-entropy loss function, an empirically tested radian sinusoidal component that is integrated with logarithm to base 2, raises the true-positive predictions. Therefore, a loss L_{sine} can be defined as:

$$L_{sine} = -\frac{1}{|Y|} \sum_{i=1}^{nC} \sin(T_i) \log_2(Y_i) \quad (5.7)$$

where $nC = 2$, and \mathbf{T} and \mathbf{Y} denote the ground-truth and network prediction, respectively, such that T_i represents the ground-truth and Y_i is the network prediction for each class i in nC . From here, computing the gradient yields:

$$\frac{\partial L_{sine}}{\partial Y_i} = -\frac{1}{|Y|} \frac{\sin(T_i)}{Y_i \log(2)} \quad (5.8)$$

The model is updated by combining the corresponding gradients of L_{sine} and L_{mhd} for optimisation.

5.3 Testing Phase

This section details the testing phase in Method 2 after a deep learning model has been successfully trained as described in Section 5.2. Every row in Figure 5.5 displays the sequential strategy where (a) a slice (2D) in a test image volume (3D) is processed through the trained deep learning model, (b) the predicted pancreas region mask is obtained, (c) max-flow and min-cuts segmentation is achieved, (d) the predicted pancreas region undergoes post-processing, and (e) the segmentation contouring (green) is highlighted against the ground-truth (red).

5.3.1 Predict Main Pancreas Region

As highlighted in the bottom methodology of Figure 5.1, a trained deep learning model performs pixel-wise prediction on a test image volume, in order to identify which region is pancreatic and otherwise. Column (a) and (b) in Figure 5.5 display four sample input slices (from four different image volumes) and their corresponding positive pancreas region (white mask) as predicted by the deep learning model. The red contouring in each image in Figure 5.5, column (b) is the ground-truth for that particular slice.

5.3.2 3D Segmentation and Refinement

Next, the predicted targeted-pancreas volume undergoes 3D based max-flow and min-cuts segmentation, coupled with effective digital contrast enhancement as described in Section 4.2 (Method 1), to reveal detailed contouring as highlighted in Figure 5.5, column (c). In Method 1, the predictions of the trained random forest may have included non-pancreatic tissue that are carried into the max-flow and min-cuts segmentation output; however, the reduced false-positive pancreas predictions in Method 2 has eliminated a significant amount of post-processing. In comparison to Method 1 (Section 4.4) the finer and more accurate prediction of the pancreas region, with less false-positive results, reduces the level of refinement employed in this method. In fact, the refinement stage merely removed scattered, false-positive voxels. Figure 5.5, column (d) highlights the max-flow and min-cuts output after masking the pancreas region and Figure 5.5,

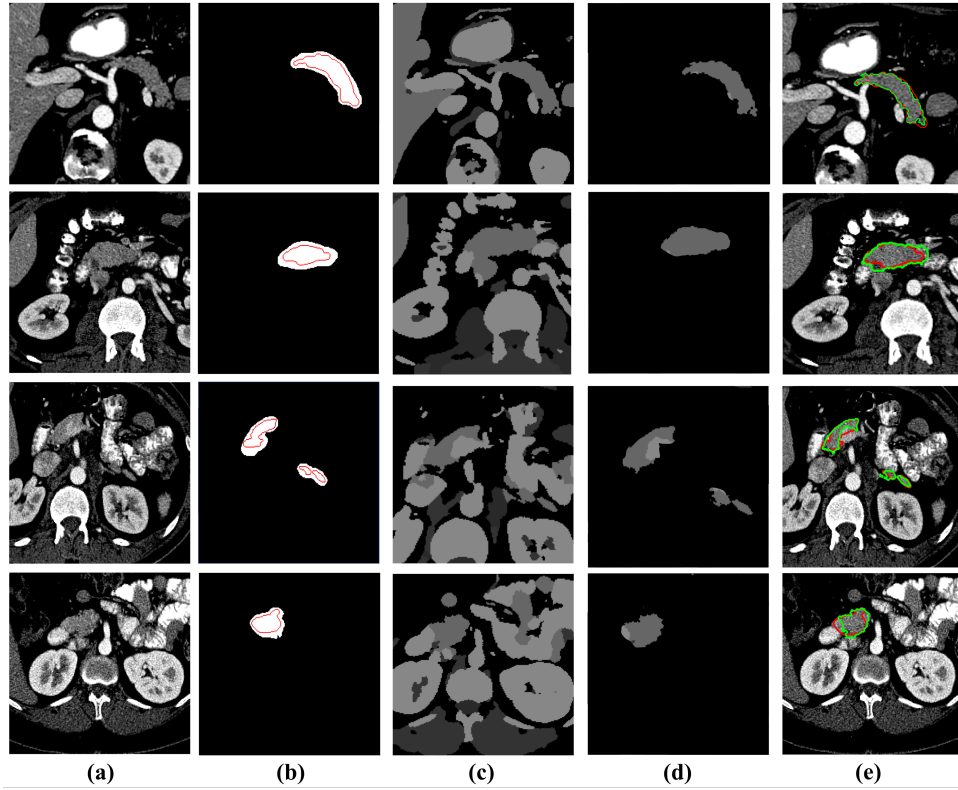


Figure 5.5: Method 2 testing phase. Column (a) displays four different slices in four different and distinct image volumes that are fed into the trained deep learning model. Column (b) reveals the predicted pancreas region mask for each corresponding input slice. Column (c) achieves max-flow and min-cuts segmentation for each input slice. Column (d) displays the pancreas region after the mask and post-processing has been applied. Column (e) reveals the segmentation contouring (green) against the ground-truth (red).

column (e) reveals the final segmentation contouring (green) against the ground-truth contouring (red).

5.4 Transitioning from Deep Learning in 2D to 3D

In 2D based deep learning approaches, the 3D radiological scans or image volumes are processed slice (2D) by slice, whereas in 3D based deep learning methods use volumetric information (as opposed to pixel information in 2D). In order to compare the performance between these two highly popular and yet inter-competitive architectures, an approach termed 3D-Method 2 is proposed. Similar to Method 2, this particular approach consists of a training stage and a testing stage as highlighted in Figure 5.6.

The first part of the training stage develops a model, defined as Rb-UNet, which aims to localise the pancreas. This model is based on the residual blocks [178] described in Section 3.2.3 (Chapter 3) which has the advantage of alleviating the vanishing gradient

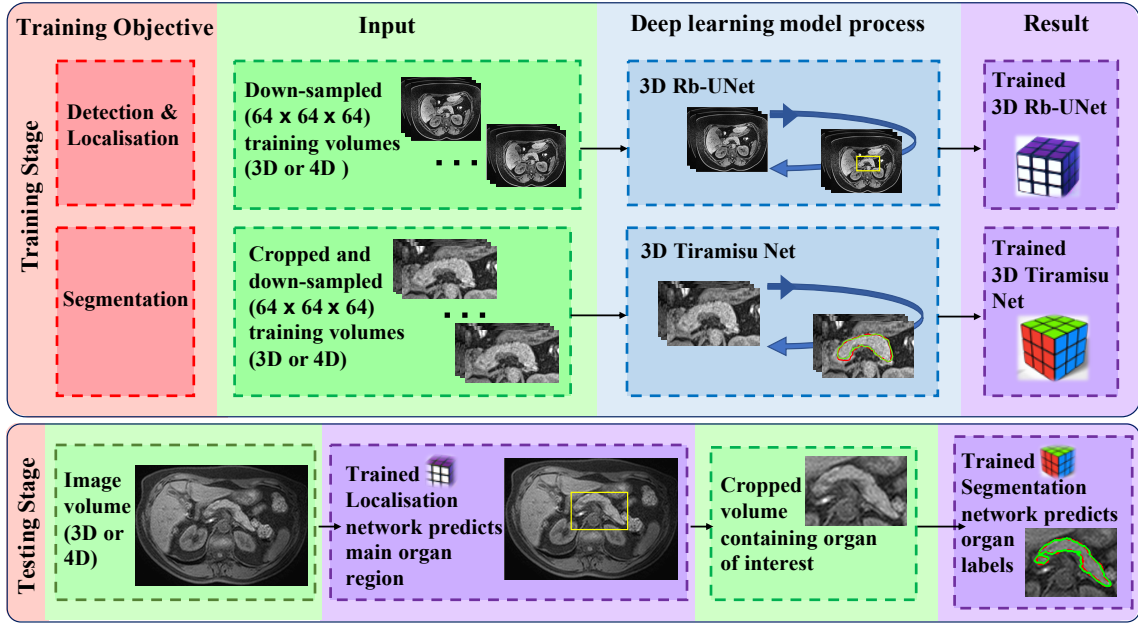


Figure 5.6: Overview of 3D-Method 2. The training stage simultaneously develops a network (3D Rb-UNet) for localising the pancreas, and a segmentation network (3D Tiramisu) to predict the labels that correspond to “pancreas” and “non-pancreas” tissue. The testing stage processes an original 3D or 4D image volume, performs a coarse segmentation to generate a bounding box capturing the main pancreas region and then processes the cropped image volume to predict the labels of that organ.

problem. The residual connections are added at each block of a baseline 3D U-Net architecture, connecting the input of convolutional layers at each scale to the outputs of the corresponding layer. Consequently, this bypass with identity connections for convolutional blocks at each scale improves the optimisation of convergence. Empirically tested, the 3D Rb-UNet model performed significantly better than the traditional 3D U-Net for localising the pancreas.

The second part of the training stage develops a 3D Tiramisu model [179] using a uniformly cropped region where the pancreas is fully present, discarding background information that is unrelated to this organ. It is important to note that, unlike the 3D Rb-UNet model which employs the full spatial context of an image volume, the 3D Tiramisu model only utilises the main pancreas region captured using a minimal bounding box generated via corresponding ground-truth labels. The main task of the 3D Tiramisu model is to perform voxel-wise predictions on whether a voxel represents “pancreas” tissue (positive label) or “non-pancreas” tissue (negative label).

In recent years, the Densely Connected Convolutional Networks (DenseNets) [180] has demonstrated promising results on image classification tasks for both natural images [201, 202] and medical scans [197, 198, 199, 200]. As introduced in Section 3.2.3 (Chapter 3), DenseNets is based on the idea of providing features ‘learned’ in earlier blocks as the inputs to succeeding blocks in the network. At every layer, the feature maps from

Layer	Transition Down (TD)	Transition Up (TU)
Batch Normalization	Batch Normalization	3×3 Transposed Convolution
ReLU	ReLU	$stride = 2$
3×3 Convolution	1×1 Convolution	
Dropout $p = 0.2$	Dropout $p = 0.2$	
	2×2 Max Pooling	

Figure 5.7: Tiramisu architecture blocks of fully convolutional DenseNets include (from left to right) the layer used in the model, Transition Down (TD) and Transition Up (TU) [179].

Architecture
Input, $m = 3$
3×3 Convolution, $m = 48$
DB (4 layers) + TD, $m = 112$
DB (5 layers) + TD, $m = 192$
DB (7 layers) + TD, $m = 304$
DB (10 layers) + TD, $m = 464$
DB (12 layers) + TD, $m = 656$
DB (15 layers), $m = 896$
TU + DB (12 layers), $m = 1088$
TU + DB (10 layers), $m = 816$
TU + DB (7 layers), $m = 578$
TU + DB (5 layers), $m = 384$
TU + DB (4 layers), $m = 256$
1×1 Convolution, $m = c$
Softmax

Figure 5.8: Architecture details of the Tiramisu (FC-DenseNet103) model, which is built from 103 convolutional layers [179]. The notations are as follows: DB stands for Dense Block, TD stands for Transition Down, TU stands for Transition Up, BN stands for Batch Normalisation and m corresponds to the total number of feature maps at the end of a block. c stands for the number of classes.

preceding layers are used as inputs, together with that layer’s very own feature maps. The main advantage of DenseNets includes strengthening the feature propagation, supporting the reuse of features, significantly reducing the number of parameters, and last but not least, alleviating the vanishing gradient problem. The Tiramisu model builds upon DenseNets to work as FCNs by adding an upsampling path to compensate for the full resolution of the input. In this architecture, it is only the feature maps created by the previous dense block that are upsampled. Furthermore, the upsampled dense block combines the contextual information retained in the other dense blocks of the same resolution. Here, the higher resolution information is passed by a standard skip connection between the downsampling and the upsampling paths. Empirically tested, the advantage of the proposed upsampling path that is built from dense blocks performs better than an upsampling path with conventional operations in DenseNets or U-Net.

In the testing stage, the fully trained 3D Rb-UNet performs a coarse segmentation of the pancreas in an (unseen) image volume and then generates a respective mini-

mum bounding box. This image volume is cropped using the bounding box. Next, this cropped image volume containing the main pancreas region processes through the fully trained 3D Tiramisu model, which in return performs finer and more detailed voxel-wise predictions of “pancreas” or “non-pancreas”.

5.4.1 Training Stage

The first phase of the training stage is the development of an organ detection and localisation approach using 3D Rb-UNet, to eventually generate a bounding box over the pancreas and remove redundant contextual information.

Detection and Localisation

At the start of this training phase, the size of the image volume is reduced to $64 \times 64 \times 64$ to limit computational costs and have sufficient resolution necessary for localisation and near isotropic resolution across x , y and z dimensions. In datasets that contain 4D image volumes, the time or temporal dimension was reduced from 150 seconds (i.e. 150 image volumes in a 4D scan) to 5 seconds (i.e. 5 image volumes in a 4D scan) using Principal Component Analysis (PCA) [184]. Moreover, data augmentation is generated as image scaled translations in the range $[1, 4]$ pixels. The weighted cross entropy loss function is employed to compensate for the class imbalance presented by diverse pancreas structures:

$$W_{cross-entropy} = -\frac{1}{N} \sum_{i=1}^N w_i^c [\hat{p}_i \log p_i + (1 - \hat{p}_i) \log(1 - p_i)] \quad (5.9)$$

where N is the total number of voxels in a training image volume, p_i is the probability of voxel i belonging to the “foreground” class in each output channel and \hat{p}_i represents the true-positive ground-truth label in the corresponding input channel. The w_i^c is fixed as inversely proportional to the probability of voxel i belonging to the “foreground” class. Afterwards, softmax with weighted cross-entropy loss is used for comparing the output of the network with the desired outcome, that is the ground-truth.

Segmentation

Using the training data, the pancreas is “cropped out” using the bounding boxes generated from the ground-truth manual labels and the image volume dimensions are reduced to $64 \times 64 \times 64$. This input data (with the temporal dimension reduced to 5 if using 4D volume data) is fed into the 3D Tiramisu network for training the main segmentation model using the same weighted cross-entropy loss as in Equation 5.9.

5.4.2 Testing Stage

At the detection and localisation stage, the fully trained 3D Rb-UNet performs a coarse segmentation, i.e. voxel-based prediction using an unseen (test) image volume. As a side note to the generalisability of this approach, for organs of interest consisting of two distinct parts (e.g. left and right kidney or left and right iliopsoas muscle) there are three classes at this stage: “left”, “right” and “background”. Otherwise, there are two main classes of “foreground” and “background” referencing the positive label prediction and negative label prediction (i.e. “pancreas” versus “non-pancreas” or “liver” versus “non-liver”). Afterwards, the image volume resamples to its original size and one or two bounding boxes are generated to “crop out” the organ of interest throughout the entire 3D or 4D volume.

The cropped test volume is fitted to $64 \times 64 \times 64$ dimensionality and processed through the fully trained 3D Tiramisu model, which performs detailed voxel-wise predictions on whether each voxel corresponds to the organ of interest or otherwise. Afterwards, the predicted organ binary mask is resampled to its original size and inserted into the corresponding spatial position of the primary input image volume.

It is noted that the identity shortcuts of the residual blocks in 3D Rb-UNet allow faster training and improved convergence than the 3D Tiramisu, which is excellent for localisation. In contrast, the 3D Tiramisu provides the advantage of higher capacity with multi-layer feature concatenation and achieves very detailed boundary-preserving fine segmentation given a localised organ of interest as the primary input.

In order to compare and discuss the differences between 3D-Method 2 and Method 2, Chapter 6 provides a sample of visual results depicting the pancreas in Figures 6.14 and 6.15, Section 6.4.4. Moreover, one of the advantages of this approach includes the usage of 4D data (3D data with a temporal dimension). As a side note to the generalisability of this approach, Chapter 8 provide a similar set of results for the kidneys in Section 8.6, Figures 8.11, 8.12, 8.13, 8.14. Furthermore, Section 8.4, Figure 8.6 and Section 8.5, Figure 8.8 provide a similar set of results for the liver and iliopsoas muscles, respectively.

5.5 Conclusion

This chapter has presented a deep learning-based approach for automatic pancreas segmentation in 3D abdominal computed tomography (CT), and Magnetic resonance imaging (MRI) scans. The proposed method builds upon Method 1 (Morphological and Multi-level Geometrical Descriptor) described in Chapter 4. A training phase firstly identifies the major pancreas region in a training data of image volumes. The major pancreas region is detected using a trained random forest that makes probability-wise

predictions of superpixels of each slice in each image volume. From here, this major region of interest feeds into a deep learning model built upon an encoder-decoder type architecture. Using a novel Hausdorff metric and sinusoidal-based loss function, the deep learning model is optimised to recognise and predict pancreatic features in a test image volume. The test phase begins by making pixel-wise predictions using the trained deep learning model; it is noted that this phase eliminates the necessity of incorporating a pancreas recognition model. This stage, followed by digital contrast enhancement and 3D max-flow and min-cuts segmentation, produces very detailed contouring of the pancreas and heavy boundary-preservation. The voxel-based algorithm of max-flow addresses the intensity consistency problem that is often the case when segmenting image volumes on a slice-by-slice basis. In this chapter, an extension of the idea in Method 2 is also proposed: 3D-Method 2 is a deep learning-based approach that employs 3D or voxel-based processing as opposed to 2D or pixel-based image processing. This method serves to demonstrate the segmentation quality of voxel-based versus pixel-based training. Chapter 6 details the experimental results that interestingly enough, highlight both the advantages and disadvantages of utilising 3D-Method 2 as opposed to Method 2 for optimal quantitative accuracy performance.

Chapter 6

Experimental Tools, Results and Analysis

6.1 Introduction

One of the most critical evaluations of different approaches includes identifying comparable advantages and disadvantages. For example, a segmentation approach that is universally applicable to different MRI sequences could produce results that surpass the state-of-the-art in mean quantitative spatial accuracy. On a different note, a segmentation technique tailored to a specific sequence could generate contouring detail that provides significant guidance towards directly stratifying an organ in terms of clinically defined normality or abnormality.

This chapter presents, discusses and compares the segmentation evaluation results that were obtained from Method 1 (Chapter 4), Method 2 and 3D-Method 2 (Chapter 5). The proposed approaches employ two MRI datasets relating to two different sequences and a publicly available CT dataset. The motivation for this chapter is to highlight key attributes of the proposed methods and address areas for potential optimisation. Section 6.2 details the acquisition and spatial information of three different and distinct datasets employed for training and testing purposes. Section 6.3 details a list of key technical tools, libraries and frameworks that are frequently used to implement deep learning algorithms for medical image segmentation. Section 6.4 details four different evaluation metrics used for computing the quantitative accuracy of the proposed methods and proceeds to perform an extensive comparison between these methods in terms of quantitative and qualitative accuracy. This section also highlights numerical implementation, including key deep learning model parameters. Section 6.5 delivers a conclusion.

6.2 Datasets

The performance of all the proposed approaches (Method 1, 2 and 3D-2) is evaluated on three pancreas datasets, all of which have been manually annotated by a senior, expert radiographer. The first two datasets contain 216 (MRI-A) and 132 (MRI-B) T2-weighted (fat suppressed) abdominal 3D MRI scans, obtained using a Philips Intera 1.5 Tesla (T) scanner and a Siemens Trio 3T scanner respectively. The CT-NIH dataset is publicly available and contains 82 abdominal contrast-enhanced CT 3D scans, acquired on Philips and Siemens MDCT scanners (120 Peak kilovoltage tube voltage)

<http://dx.doi.org/10.7937/K9/TCIA.2016.tNB1kqBU>. Please refer to Appendix A to view a set of randomly selected 2D images (slices) taken from volumes in MRI-A and MRI-B, and CT-NIH.

6.2.1 MRI

Every image volume in MRI-A consists of 50 axial slices, with each slice of spatial size 384×384 pixels and $2mm$ thickness, and $0.9766mm$ in voxel width and height. In MRI-B, every image volume consists of 80 axial slices, with each slice of spatial size 320×260 pixels and $1.6mm$ thickness, and $1.1875mm$ in voxel width and height. The subjects who underwent the MRI scan were over 18 years of age and displayed early signs of type 2 diabetes.

6.2.2 CT

The CT-NIH dataset contains image volumes of spatial size 512×512 pixels (between 181 and 466 slices) and slice thickness ranging from $1.5 - 2.5mm$. The CT scans were acquired from 53 male and 27 female subjects. A radiologist selected sixty-five subjects from patients who did not have any major abdominal pathologies nor pancreatic cancer lesions; the remaining seventeen subjects were healthy kidney donors scanned prior to nephrectomy. The subjects' age range is 18 - 76 years with a mean age of 46.8 ± 16.7 .

6.3 Technical Tools, Libraries and Frameworks

This section provides a summary of key technical tools, libraries and frameworks that are frequently used to implement deep learning algorithms for medical image processing, in particular medical image segmentation. Both approaches described in Chapters 4 and 5 are implemented using Matlab (Releases 2016b - 2019b). 3D-Method 2 implemented in Python 3.0 (Keras).

6.3.1 MATLAB

MATLAB is an interactive computing environment and high-level programming language developed by MathWorks [143]. MATLAB performs matrix manipulations and data plotting, and interfacing with programs written in other languages such as C, C++ and Python. MATLAB provides an Image Processing Toolbox to solve problems and explore ideas for medical image analysis [144]. This toolbox provides a wide range of image processing functions for converting, enhancing and analysing digital images. A number of these functionalities include spatial image transformations, morphological operations such as edge detection and noise removal, and ROI processing and filtering. In addition to the Image Processing Toolbox, there are three further essential toolboxes that have enabled successful implementation of image segmentation methodologies: Computer Vision System Toolbox, Statistics and Machine Learning Toolbox and the Deep Learning Toolbox.

Computer Vision System Toolbox

The Computer Vision System Toolbox provides algorithms and functions for designing and simulating computer vision, and video processing systems. It is possible to perform feature detection, extraction and matching, as well as perform object detection and tracking. For 3D computer vision, the system toolbox supports 3D reconstruction, as well as 3D point cloud processing. Algorithms for deep learning and machine learning can enable the programmer to implement face detection algorithms and other everyday objects using pre-trained detectors.

Statistics and Machine Learning Toolbox

The Statistics and Machine Learning Toolbox provides functionalities to describe, analyse and model data. It is possible to use descriptive statistics to plot exploratory data analysis, fit probability distributions to data, generate random numbers for Monte Carlo simulations, and perform hypothesis tests. This toolbox provides supervised and unsupervised machine learning algorithms, including support vector machines (SVMs), boosted and bagged decision trees, k -nearest neighbour, k -means, k -medoids, hierarchical clustering, Gaussian mixture models and hidden Markov models. Furthermore, the Statistics and Machine Learning Toolbox provides tools that perform regression and classification algorithms in order to draw inferences from data and build predictive models. When using higher-dimensional data analysis, the Statistics and Machine Learning Toolbox provides feature selection, stepwise regression, principal component analysis (PCA) and regularisation methods, all of which allows the programmer to identify variables or features that impact the model being implemented.

Deep Learning Toolbox

The Deep Learning Toolbox, formerly known as the Neural Network Toolbox, provides a framework for designing and implementing deep neural networks with algorithms and pre-trained models. It is possible to use convolutional neural networks (ConvNets, CNNs) and long short-term memory (LSTM) networks to perform classification and regression on image, time-series, and text data. Furthermore, several apps and plots enable programmers to visualise activations, fine-tune network architectures and monitor training progress. It is also possible to utilise many pre-trained (popular) networks to fine-tune including GoogLeNet, VGG-16, VGG-19 and AlexNet. It is also possible to create, modify and analyse complex CNN architectures using MATLAB visualisation tools. MATLAB can also bring together multiple domains in a single workflow; for example, it is possible to implement tools and functions for deep learning, and also for a variety of domains that feed into deep learning algorithms, e.g., signal processing, computer vision, and data analytics. Currently, there are three computational methods to train a CNN on MATLAB: CPU-based¹, GPU-based², and cloud-based GPU³. A CPU-based computation is the simplest and most readily available option but is recommended only for simple examples using a pre-trained network. Using a GPU reduces network training time from days to hours, and multiple GPUs can further accelerate processing and reduce computational time. In a cloud-based GPU computation, the MATLAB code can be executed using cloud resources.

The implementation of deep learning architectures, coupled with modular tools and functions for managing large data sets, makes MATLAB a useful tool for heuristic network training and evaluation. Not only does MATLAB offer a platform to build deep learning models from scratch, but it is also possible to import pre-trained models to visualise and debug intermediate results after adjusting training parameters. This software application can be used to learn about and implement deep learning architectures in a theoretical, analytical and practical manner. Furthermore, this computing environment also enables programmers to label objects within images interactively and can automate ground-truth labelling within image volumes for training and testing deep learning models.

It is also possible to exchange deep learning models with other popular open-source software libraries, such as TensorFlow and import models from TensorFlow-Keras.

¹The central processing unit (CPU) is the unit which performs most of the processing inside a computer.

²A graphics processing unit (GPU) is a computer chip that performs rapid mathematical calculations, mainly rendering images.

³A cloud is network of remote servers hosted on the Internet to store, manage, and process data, rather than a local server or a personal computer.

6.3.2 TensorFlow and Keras

TensorFlow [146] is an open source software library for high performance numerical computation; it has a flexible architecture that allows deployment of computation across CPU, GPU and TPU⁴ platforms. Keras [153], on the other hand, is a high-level neural networks application programming interface (API), written in Python and capable of running on top of TensorFlow. It was developed to enable fast experimentation. Keras offers a deep learning library that allows easy and fast prototyping that is user-friendly; supports both CNNs and recurrent networks, and runs on CPU and GPU. The user-friendliness offers consistent and simple APIs, and thus minimises the number of user actions required for frequent use, providing clear and actionable feedback upon user error. Keras also has modularity: fully-configurable modules can be employed, particularly CNN layers, loss functions, initialisation schemes, activation functions and regularisation schemes, which can also be combined to create new models.

6.4 Evaluation

The performance of the proposed approaches is evaluated using the Dice similarity coefficient (DSC) and Jaccard index (JI) method, in addition to precision (PC) and recall (RC) [69]. The JI is a measure of the spatial overlap between the segmentation outcome and desired outcome (i.e. the ground-truth or expert-led manual annotation of the organ). Unlike the JI metric, which only counts true-positives (i.e. positive labels common to segmentation and ground-truth) once in both the numerator and denominator, the DSC function does not possess the property of triangle inequality and this particular metric function retains sensitivity in more dissimilar data, and assigns less weight to outliers [175].

If G represents the volumetric ground-truth annotation and S represents the corresponding automatic segmentation result, then the DSC accuracy percentage of S relative to G can be defined as: $DSC = \left(\frac{2|G \cap S|}{|G| + |S|} \right) \times 100$. Similarly, the JI accuracy percentage of S relative to G can be defined as: $JI = \left(\frac{|G \cap S|}{|G \cup S|} \right) \times 100$. The precision normalises the correct segmented region, $S \cap G$, against the segmentation, S : $Precision = \left(\frac{|S \cap G|}{|S|} \right) \times 100$. The recall (i.e. sensitivity) normalises $S \cap G$ against the ground-truth, G and can be defined as: $Recall = \left(\frac{|S \cap G|}{|G|} \right) \times 100$.

⁴A tensor processing unit (TPU) is an AI application-specific integrated circuit (ASIC) for machine learning implementation.

6.4.1 Method 1: Morphological and Multi-level Geometrical Descriptor Analysis

The training and testing image volumes have been randomly split into 196 and 20 for MRI-A, 112 and 20 for MRI-B, and 62 and 20 for CT-NIH. The defined parameters for the maximal-flow and minimum graph-cuts approach described in Section 4.3 are as follows: step-size of the augmented Lagrangian method is 0.25; error bound for convergence is 0.0001; maximum number of iterations is 400; and the step-size for the gradient-projection step to the total-variation function is 0.03.

For each experiment in Section 4.4.1, the values for N_{r_1} , N_{r_2} , N_{r_3} and N_{r_4} are 32, 52, 64 and 16 respectively. The group constraint error threshold, $t_m = 0.15$ and the contour curvature threshold, $t_c = 0.27$. In Section 4.4.2, the value of n (set of distances) is 6, and in Section 4.4.3, the value of the mean error threshold, $t_p = 0.15$ and the chosen number of groups in a set based on k -medoids clustering, n is set to 16, 48, 36 and 12 for slice groups R_1 , R_2 , R_3 and R_4 respectively.

6.4.2 Method 2: Integrating Deep Learning

The optimisation algorithm used for training is SGD, with momentum 0.9 and an initial learning rate of 0.001. The learning rate drop period is 50 and the learning rate drop factor is 0.5. The maximum number of epochs is 320 and the size of the mini-batch to use for each training iteration is set to 10. The training and testing image volumes have been randomly split into 196 and 20 for MRI-A, 112 and 20 for MRI-B, and 62 and 20 for CT-NIH. The experiments conducted using MRI-A, MRI-B and CT-NIH employ 10-fold, 6-fold and 4-fold cross-validation, respectively.

Analysis of Hausdorff-Sine Loss:

Figures 6.1 and 6.2 compares the quantitative segmentation results for MRI (in DSC) using loss functions Hausdorff-Sine, Hausdorff, Cross-entropy, Dice [207] and Jaccard [206] across thresholds in range [0.05, 0.95]. The loss Hausdorff-Sine achieves the highest mean DSC, irrespective of the segmentation threshold; in fact, the accuracy in DSC increases with a higher threshold. Across differing threshold probabilities, the Cross-entropy has a drawback on predicting true-positive (“pancreas”) pixels, and therefore forces the resulting “optimum” probability to remain at approximately 0.5. Although the Dice loss minimises the class distribution distance, squaring the weights in the backpropagation stage causes instability and a higher rate of false-negative predictions. Similarly, the Jaccard loss suffers from low true-positive predictions. Empirically tested, the Hausdorff loss minimises the maximum deviation between a prediction and desired outcome; however, the addition of a sinusoidal component increases non-linearity during

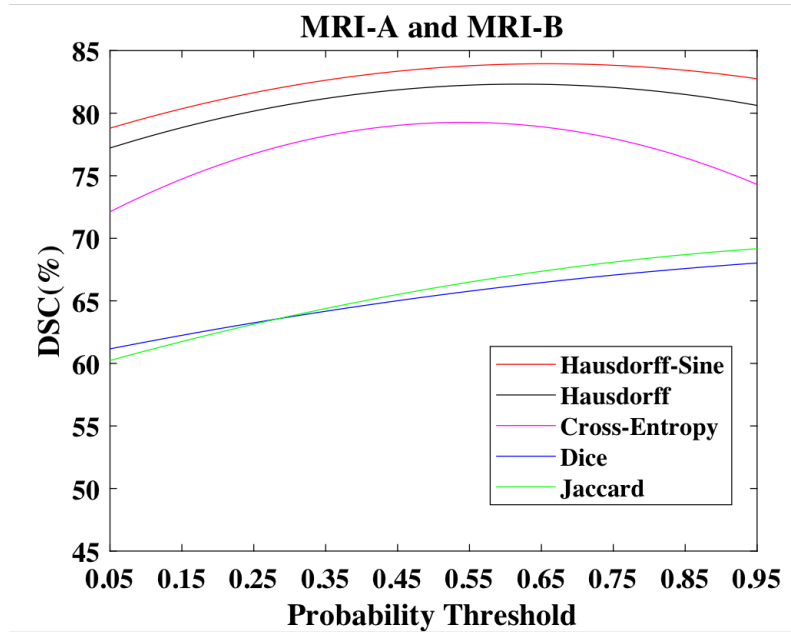


Figure 6.1: Overview of MRI segmentation accuracy results (in DSC) of models that are trained with different loss functions. The proposed Hausdorff-Sine Loss performs better across thresholds in range $[0.05, 0.95]$ in comparison to the conventional cross-entropy and other loss functions.

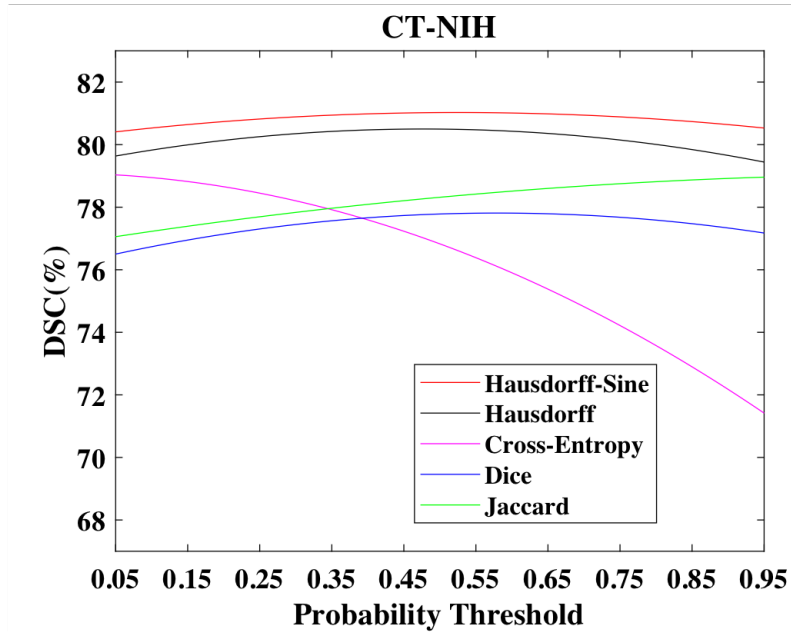


Figure 6.2: Overview of CT-NIH segmentation accuracy results (in DSC) of models that are trained with different loss functions. The proposed Hausdorff-Sine Loss performs better across thresholds in range $[0.05, 0.95]$ in comparison to the conventional cross-entropy and other loss functions.

training, and thus Hausdorff-Sine achieves a better prediction of true-positive pixels across differing probability thresholds, whilst delivering strong discrimination of “non-pancreas” pixels.

The mean DSC and standard deviation using Cross-entropy is $83.0\% \pm 8.0\%$ (MRI-A), $83.9\% \pm 3.1\%$ (MRI-B) and $78.2\% \pm 8.8\%$ (CT-NIH). However, employing Hausdorff-Sine raises the mean DSC by approximately 1% (MRI-A), 2% (MRI-B) and 5% (CT-NIH) and improves statistical stability by reducing the mean standard deviation by approximately 3% (MRI-A), 1% (MRI-B) and 4% (CT-NIH). The Hausdorff distance (HSD) metric, presented as mean \pm standard deviation, represents the maximum deviation between two surfaces and is very sensitive to outliers. In contrast to Cross-entropy that delivers $12.71 \pm 8.62mm$ for MRI-A, the novel loss function Hausdorff-Sine achieves less deviation from the ground-truth contouring with $10.55 \pm 3.72mm$. Similarly, Cross-entropy delivers $22.41 \pm 6.59mm$ for MRI-B whereas Hausdorff-Sine achieves a significantly more robust $9.08 \pm 1.98mm$. Using CT-NIH, Cross-entropy delivers $13.30 \pm 5.93mm$ whereas Hausdorff-Sine achieves $10.34 \pm 3.39mm$. The corresponding ROC curves are illustrated in Figures 6.3 and 6.4, from where the area under the ROC curve and true-positive rate can be used as performance metrics. The inferior performance of conventional loss functions demonstrate that these loss functions cannot be optimal in extremely unbalanced segmentation for small anatomical structures, whereas Hausdorff-Sine generally improves the true-positive accuracy results.

6.4.3 Comparison between Method 1 and 2

Table 6.1 displays the DSC, JI, PC, RC and Hausdorff distance [80] results using Method 1 and Method 2 for datasets, MRI-A, MRI-B and CT-NIH as mean \pm standard deviation [lowest, highest]. Figures 6.7 and 6.8 displays the final segmentation results for Method 1 and 2 in six MRI scans, equally split between MRI-A and MRI-B. Despite columns (a), (b) and (c) being part of MRI-A, notice the variation between intensity and image contrast in the original axial MRI slices. Columns (d), (e) and (f) corresponds to exemplars from dataset MRI-B. In particular, the contouring of the pancreas can appear less well-defined with blurred boundaries between the organ and surrounding tissue, and therefore affect the overall segmentation accuracy. Similarly, Figures 6.5 and 6.6 displays the visualisations of four different pancreas segmentation results in the CT-NIH dataset. There are apparent, evident differences in the shape and structure of slice-by-slice pancreatic contours as well as the overall 3D reconstruction.

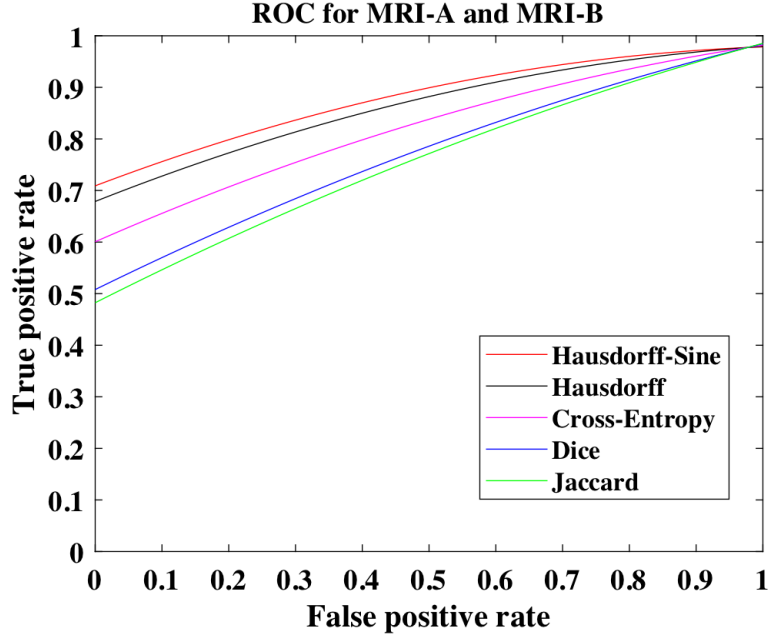


Figure 6.3: ROC curves averaged on MRI-A and MRI-B using the deep learning model in Method 2 trained via the proposed Hausdorff-Sine loss, and other loss functions including Hausdorff (alone), Cross-entropy, Dice and Jaccard.

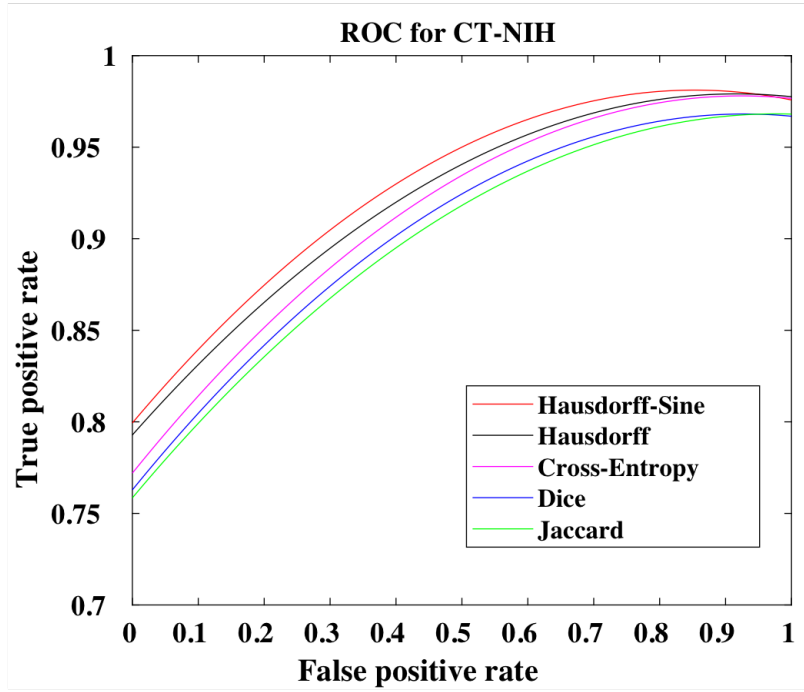


Figure 6.4: ROC curves averaged on CT-NIH using the deep learning model in Method 2 trained via the proposed Hausdorff-Sine loss, and other loss functions including Hausdorff (alone), Cross-entropy, Dice and Jaccard.

Data	Method	DSC (%)	JI (%)	PC (%)	RC (%)	Haus. (mm)
MRI-A	1	79.6 \pm 5.7 [68.6, 87.5]	66.5 \pm 7.9 [52.4, 77.8]	85.7 \pm 11.5 [71.5, 98.9]	94.7 \pm 10.7 [50.5, 99.7]	13.59 \pm 4.34 [6.12, 20.36]
	2	84.1 \pm 4.6 [72.1, 89.6]	72.9 \pm 6.5 [56.4, 81.1]	95.5 \pm 6.3 [71.7, 99.7]	97.6 \pm 3.0 [89.9, 100.0]	10.55 \pm 3.72 [6.06, 17.81]
	3D-2	89.9 \pm 3.4 [82.4, 94.2]	81.87 \pm 5.6 [73.8, 89.1]	87.5 \pm 3.8 [78.9, 91.7]	92.5 \pm 3.6 [84.1, 9.7]	10.23 \pm 5.98 [3.18, 20.62]
MRI-B	1	81.6 \pm 5.1 [71.3, 88.0]	69.2 \pm 7.1 [55.5, 78.6]	82.2 \pm 5.8 [73.1, 93.5]	84.8 \pm 8.3 [69.4, 96.0]	14.28 \pm 4.02 [8.14, 22.51]
	2	85.7 \pm 2.3 [79.9, 90.3]	75.1 \pm 3.5 [66.5, 82.4]	96.1 \pm 3.6 [86.7, 100.0]	99.3 \pm 0.7 [99.9, 100.0]	9.08 \pm 1.98 [4.91, 14.80]
	3D-2	90.2 \pm 5.1 [72.2, 95.8]	82.6 \pm 7.8 [56.4, 91.9]	88.0 \pm 6.9 [58.4, 95.7]	92.9 \pm 4.2 [81.7, 97.1]	7.43 \pm 5.26 [2.77, 20.22]
CT-NIH	1	79.3 \pm 4.4 [72.8, 86.0]	66.1 \pm 6.2 [58.2, 75.5]	88.6 \pm 6.9 [74.5, 99.5]	97.6 \pm 1.9 [94.7, 99.8]	12.52 \pm 1.98 [9.15, 15.45]
	2	83.1 \pm 5.3 [69.3, 89.6]	71.4 \pm 7.4 [53.1, 81.1]	97.6 \pm 5.3 [75.2, 99.9]	97.1 \pm 3.7 [84.9, 100.0]	10.34 \pm 3.39 [6.26, 19.44]
	3D-2	84.7 \pm 7.9 [65.4, 95.1]	74.2 \pm 11.4 [48.7, 86.1]	87.2 \pm 5.6 [71.7, 95.3]	82.6 \pm 10.3 [60.2, 94.9]	9.56 \pm 5.67 [3.30, 19.91]

Table 6.1: Comparative summary of DSC, JI, PC, RC and Hausdorff (Haus.) results shown as mean \pm standard deviation [lowest, highest] in MRI-A, MRI-B and CT-NIH using Methods 1, 2 and 3D-2.

Pancreas Boundary Accuracy

In Method 1, the Hausdorff distance (HSD) metric is presented as mean \pm standard deviation in CT-NIH, MRI-A and MRI-B as $12.52 \pm 1.98mm$, $13.59 \pm 4.34mm$ and $14.28 \pm 4.02mm$, respectively. This metric, which represents the maximum deviation between two surfaces and is very sensitive to outliers, indicates that the approach is slightly more consistent for CT data. The relatively lower HSD in CT implies that slices have greater co-dependency, and therefore the proposed approach delivers a stable performance in predicting between successive slices. In contrast, the higher HSD value in MRI reflects less co-dependency between successive slices and higher shape variability. In such cases, the major pancreas region included excess background tissue that formed part of the final segmentation result, or that pancreas tissue (mostly comprising part of the tail or body) was eliminated.

In contrast to Method 1, the mean HSD in MRI-A and MRI-B using Method 2 is relatively lower as in $10.55 \pm 3.72mm$ and $9.08 \pm 1.98mm$, respectively. The significant drop in mean value of approximately $5mm$ in MRI-B demonstrates the effectiveness of the deep learning model; the pixel-wise predictions of “non-pancreatic” tissue is comparably less than Method 1, which solely relies on the superpixel (size 32) predictions of

a trained random forest. MRI-A achieves a relatively modest reduction in mean HSD of approximately $2mm$, mostly reflective of the greyscale intensity distribution diversity in this particular dataset. Interestingly, Method 2 also improves the mean HSD achieved in CT by reducing the value by approximately $2mm$. A slight increase in standard deviation indicates that, although the overall contouring accuracy has improved, some instances of under-segmentation can be addressed through network optimisation (discussed in Chapter 7, Section 7.3).

Statistical Graph Comparison

In Method 1, a box and whisker plot representation for all three datasets is displayed in Figure 6.9. The standard deviation is a reflection of intensity variation of surrounding tissue coming into contact with the final, resultant pancreas segmentation. In CT-NIH and MRI-B, 60% of all segmentation results score 80% or above in DSC, compared to 45% of all segmentation results in MRI-A. This difference in accuracy scores highlights the performance of the approach concerning image quality and distribution of greyscale intensity. The broader range between the median and upper quartile for MRI-A, in comparison to MRI-B, suggests a higher degree of variation in the test volumes' shape, size and level of coarse image noise. Thus, a smaller interquartile range for MRI-B indicates a similar degree of error within that dataset affects all resultant segmentations. The relatively lower standard deviation in CT-NIH reflects the lesser degree of variation between image volumes as opposed to MRI-A and MRI-B that suffer from higher image noise and artefacts.

Method 2 raises the overall accuracy scores in DSC, such that 85% of all segmentations score above 80% in MRI-A (versus 45% using Method 1), 95% of all segmentations score above 80% in MRI-B (versus 60% using Method 1) and 75% of all segmentations score above 80% in CT-NIH (versus 45% using Method 1). Although there are relatively more outliers present in comparison to Method 1, as highlighted in Figure 6.10, the mean DSC standard deviation is almost halved for MRI-B using Method 2 and reduced for MRI-A. The significantly different range between the median and maximum score in MRI-A highlights that 75% of all results (ranging from 84.39% – 89.56%) are closer to the mean DSC using Method 2, as opposed to 55% of all results (ranging from 79.14% – 87.53%) using Method 1. Similarly, 65% of MRI-B DSC results (ranging from 85.16% – 90.33%) are closer to the mean score in Method 2, as opposed to 55% of all results (ranging from 82.51% – 88.03%) using Method 1. Using the CT-NIH dataset, 70% of all DSC results (ranging from 82.59% – 89.56%) are closer to the mean score in Method 2, as opposed to 65% of all results (ranging from 78.14% – 86.03%) using Method 1.

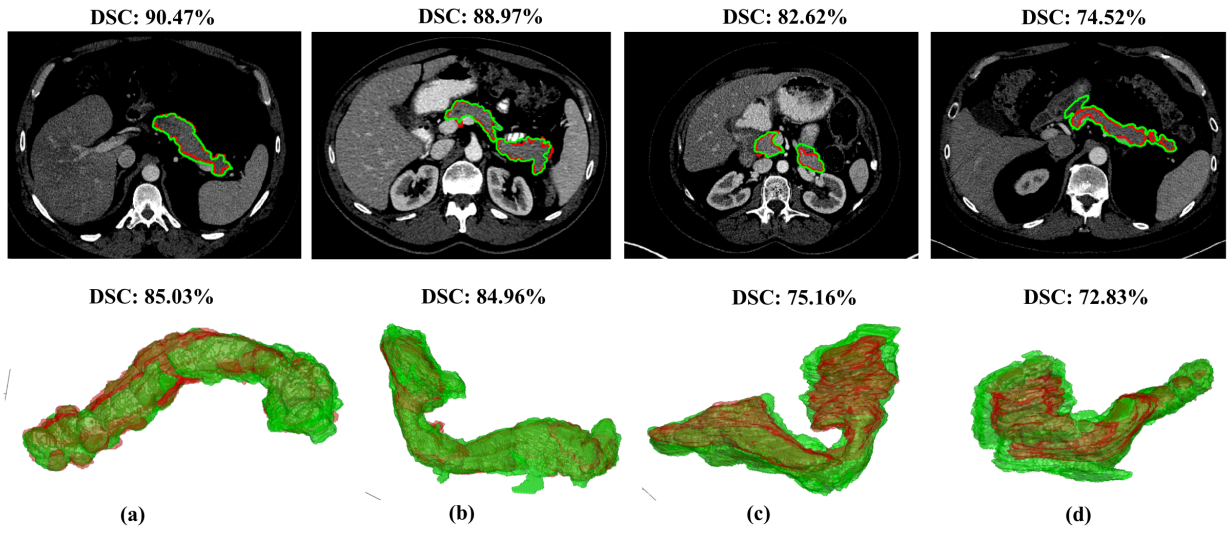


Figure 6.5: Segmentation results in four different CT image scans (volumes) using Method 1. Every column corresponds to a single CT volume. From left, first row displays sample CT axial slices with segmentation outcome (green) against ground-truth (red), and computed DSC; second row displays 3D reconstruction of entire pancreas (green) segmentation against its ground-truth (red) with computed DSC.

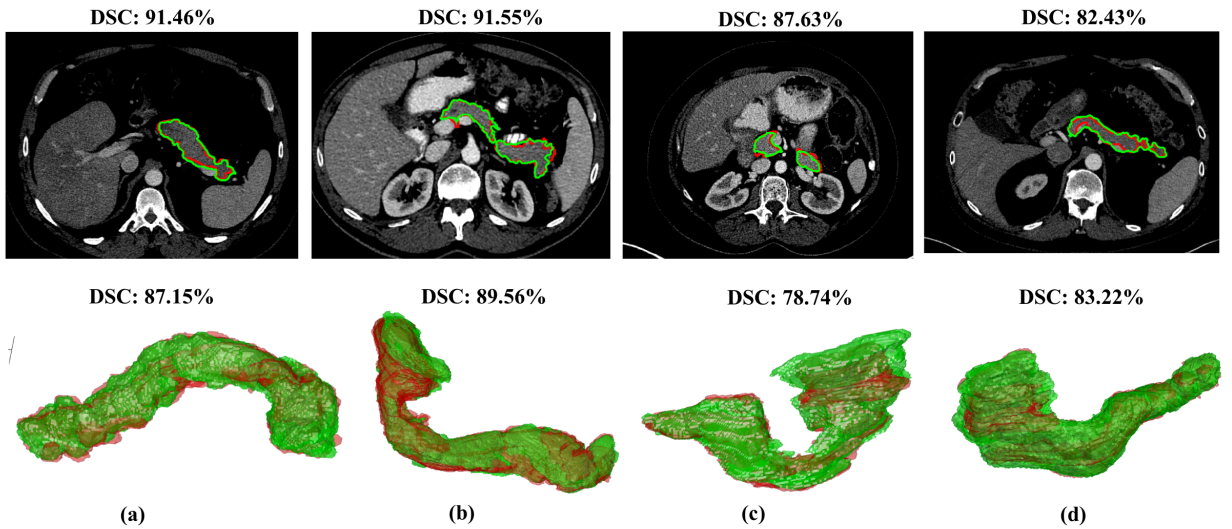


Figure 6.6: Segmentation results in four different CT image scans (volumes) using Method 2. Every column corresponds to a single CT volume. From left, first row displays sample CT axial slices with segmentation outcome (green) against ground-truth (red), and computed DSC; second row displays 3D reconstruction of entire pancreas (green) segmentation against its ground-truth (red) with computed DSC.

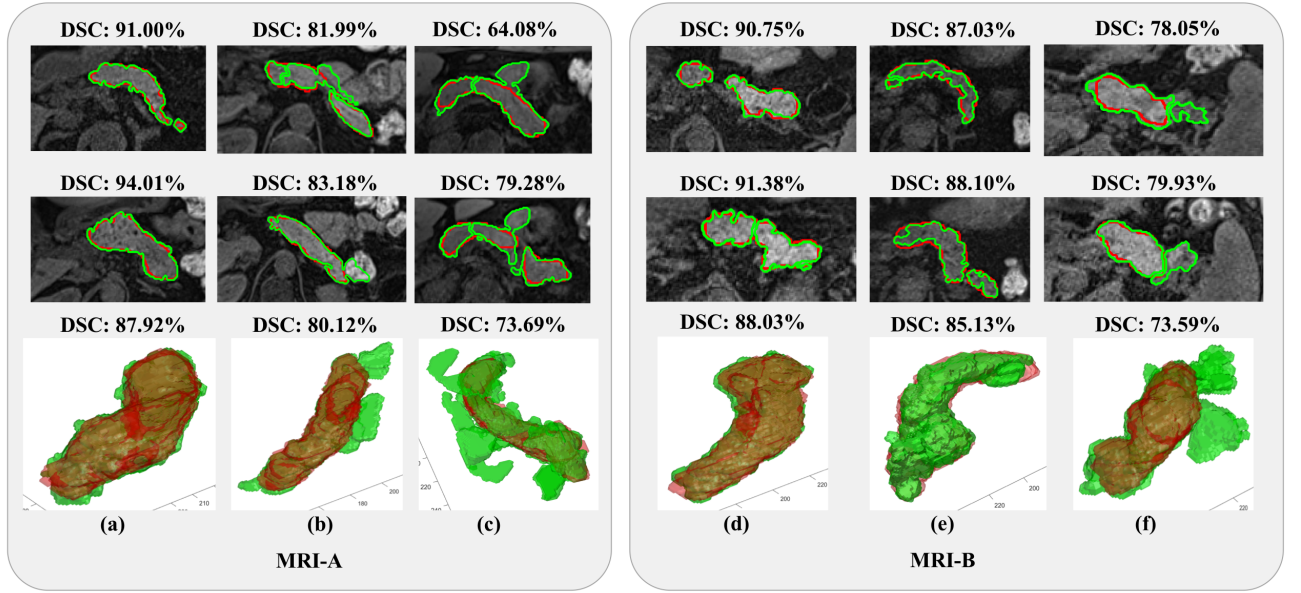


Figure 6.7: Segmentation results in six different MRI scans (volumes) for Method 1 (Morphological and Multi-level Geometrical Descriptor Analysis). Every column corresponds to a single MRI volume. From left, first row displays sample MRI axial slices with segmentation outcome (green) against ground-truth (red), and computed DSC; second row displays 3D reconstruction of entire pancreas (green) segmentation against its ground-truth (red) with computed DSC.

Precision and Recall

In Method 1, the precision and recall results in CT-NIH significantly outperform state-of-the-art approaches reported in [1, 21, 29, 78]. Therefore, comparatively high precision and high recall scores reflect a high true-positive rate and a low false-negative rate. As shown in Figure 6.11, MRI-A and MRI-B report strong precision and recall results, but the differences in standard deviation reflect higher variations of correct and incorrect “pancreas” pixel classification. In general, the initial detection of the major pancreas region guarantees that the pancreas will be segmented (albeit with varying DSC accuracy) while maintaining higher mean precision and recall scores.

Using Method 2, the results obtained for MRI-A and MRI-B boast an increase in precision and recall, as highlighted in Table 6.1. Using CT-NIH, the mean precision is raised by approximately 9% in comparison to Method 1, and a slight reduction in false-negative pixels (mean recall) is a reflection of under-segmentation for particular cases against a higher maximum recall of 100%. The reduced non-pancreatic tissue prediction achieved in Method 2 raises the mean precision and recall scores for MRI-B by approximately 12% and 15%, respectively. Similarly, Method 2 raises the mean precision score for MRI-A by approximately 10%, reflecting the significantly lower false-positive rate. Interestingly, an approximate 3% increase in mean recall demonstrates the effectiveness of Method 1 in delivering a low false-negative rate.

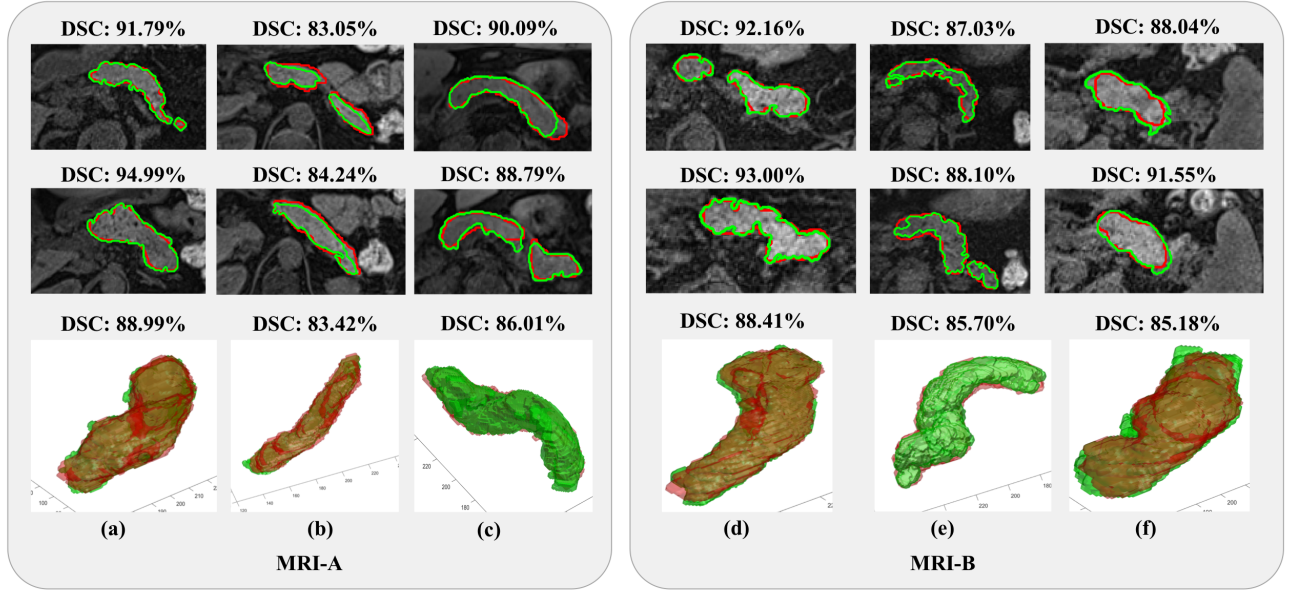


Figure 6.8: Segmentation results in six different MRI scans (volumes) for Method 2 (Deep Fusion). Every column corresponds to a single MRI volume. From left, first row displays sample MRI axial slices with segmentation outcome (green) against ground-truth (red), and computed DSC; second row displays 3D reconstruction of entire pancreas (green) segmentation against its ground-truth (red) with computed DSC.

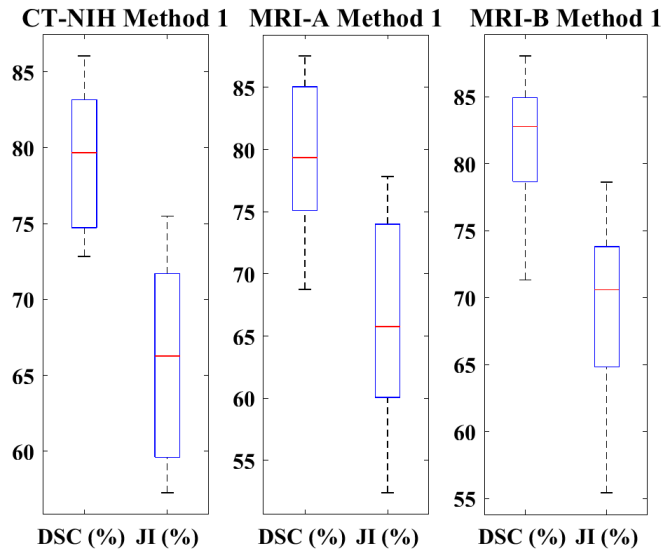


Figure 6.9: Dice similarity coefficient (DSC) and Jaccard index (JI) box plots for datasets CT-NIH, MRI-A and MRI-B using Method 1.

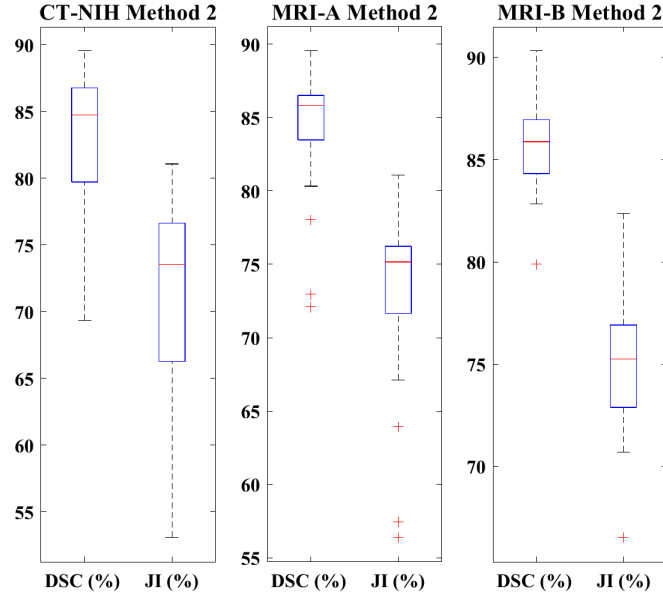


Figure 6.10: Dice similarity coefficient (DSC) and Jaccard index (JI) box plots for datasets CT-NIH, MRI-A and MRI-B using Method 2.

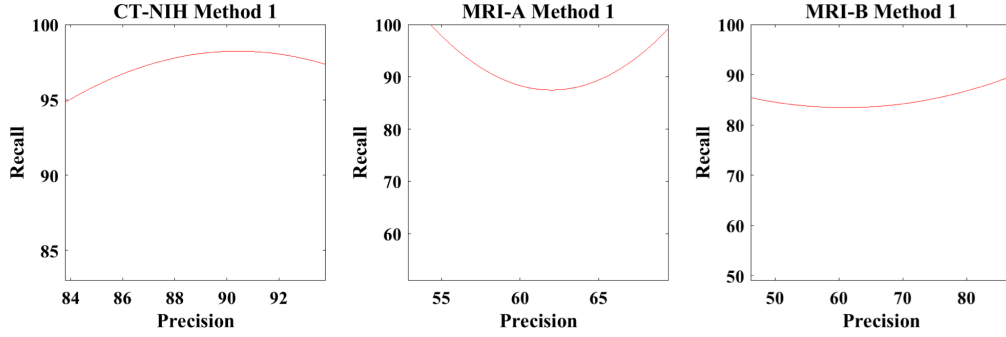


Figure 6.11: Recall against precision plots for datasets CT-NIH, MRI-A and MRI-B using Method 1.

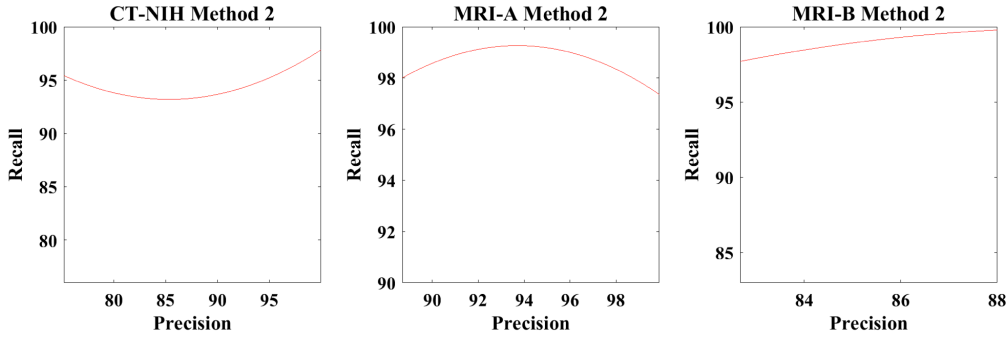


Figure 6.12: Recall against precision plots for datasets CT-NIH, MRI-A and MRI-B using Method 2.

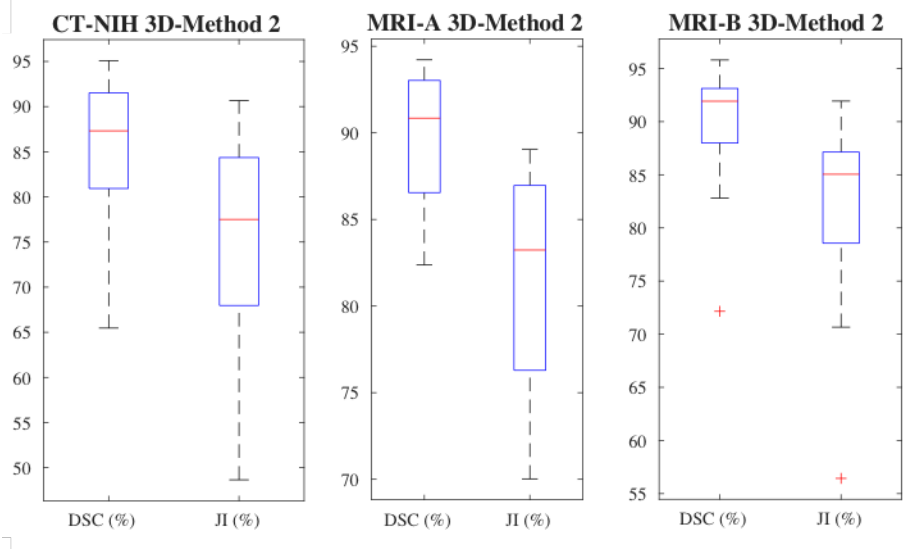


Figure 6.13: Dice similarity coefficient (DSC) and Jaccard index (JI) box plots for datasets CT-NIH, MRI-A and MRI-B using 3D-Method 2.

6.4.4 3D-Method 2

The optimisation algorithm used for training is Adam⁵ [203] with an initial learning rate of 0.0001. The hyperparameters include a reduction rate (0.8), growth rate (12), momentum (0.9), weight decay (10^{-8}) and a drop-out rate (0.2). The learning rate drop period is 50 and the learning rate drop factor is 0.5. The maximum number of epochs is 400 and the size of the mini-batch to use for each training iteration is set to 4.

6.4.5 Comparison between Method 2 and 3D-Method 2

The addition of 3D-Method 2 aims to compare and contrast the effectiveness of Method 2 in terms of quantitative accuracy and statistical stability. Table 6.1 displays the DSC, JI, PC, RC and Hausdorff (Haus.) results using Method 2 and 3D-Method 2 for all three datasets as mean \pm standard deviation [lowest, highest]. Figure 6.14 compares the visualisations of three different pancreas segmentation results in MRI-A and MRI-B with the results obtained using Method 2, and Figure 6.13 displays the corresponding box and whisker plot representation of the segmentation results using 3D-Method 2.

Interestingly, the mean DSC in CT-NIH raises from $83.1 \pm 5.3\%$ to $84.7 \pm 7.9\%$ using the 3D deep learning approach. With that said, the increase in standard deviation demonstrates that replacing 2D operations with corresponding 3D counterparts to optimise learning voxelised contextual information results in higher quantitative accuracy variation. In fact, while Method 2 boasts a strong mean precision and recall as in $97.6 \pm 5.3\%$ and $97.1 \pm 3.7\%$ respectively, 3D-Method 2 suffers from false-positive and false-negative

⁵Adam is an adaptive learning rate optimisation algorithm designed for training deep neural networks.

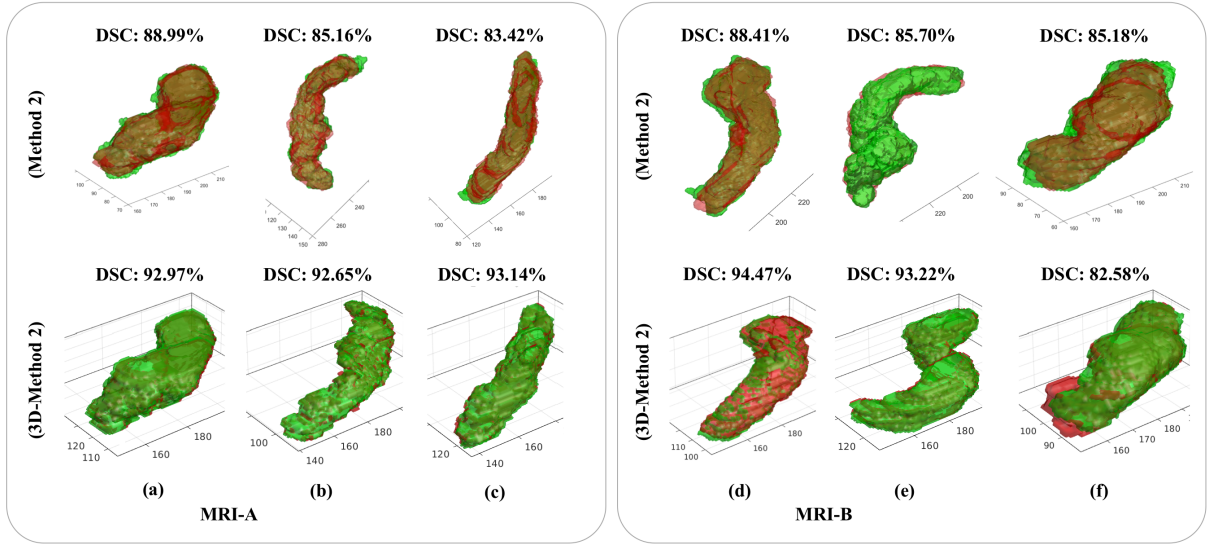


Figure 6.14: Pancreas segmentation results in six different MRI image scans (volumes) using Method 2 and 3D-Method 2. Every column of two 3D pancreas reconstructions highlights the segmentation results and DSC in the same MRI volume using Method 2 (top row) with 3D-Method 2 (bottom row). The segmentation outcome (green) overlaps the ground-truth (red) with computed DSC.

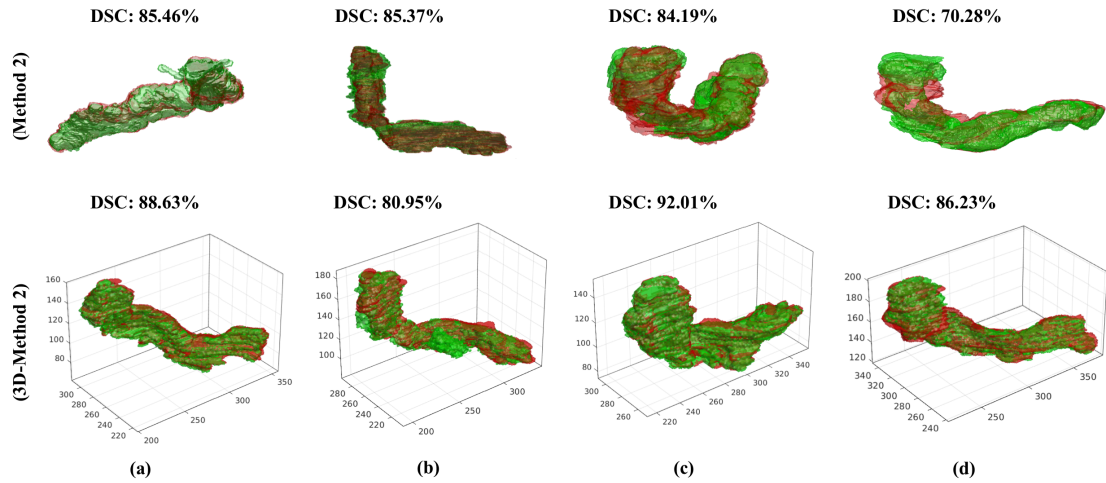


Figure 6.15: Pancreas segmentation results in four different CT image scans (volumes) using Method 2 and 3D-Method 2. Every column of two 3D pancreas reconstructions highlights the segmentation results and DSC in the same CT volume using Method 2 (top row) with 3D-Method 2 (bottom row). The segmentation outcome (green) overlaps the ground-truth (red) with computed DSC.

predictions as in $87.2 \pm 5.6\%$ and $82.6 \pm 10.3\%$. This stark contrast in recall indicates that the volumetric nature of the pancreas (object of interest) cannot be used to dictate the best architecture as 3D based. It can be argued that the 2D (slice-by-slice) image processing supports the network to learn detailed contextual information that is otherwise missed in 3D feature extraction. Figure 6.15 compares the visualisations of four different pancreas segmentation results in CT-NIH with the results obtained using Method 2, and Figure 6.13 displays the corresponding box and whisker plot representation of the segmentation results using 3D-Method 2.

The evaluation of the MRI datasets presents a similar finding in comparison to the CT dataset. Whereas Method 2 delivers a mean DSC of $84.1 \pm 4.6\%$ and $85.7 \pm 2.3\%$ for MRI-A and MRI-B respectively, 3D-Method 2 achieves significantly improved results as in $89.9 \pm 3.4\%$ and $90.2 \pm 5.1\%$. However, the mean precision differs with 3D-Method 2 achieving $87.5 \pm 3.8\%$ and $88.0 \pm 6.9\%$ for MRI-A and MRI-B respectively, whereas Method 2 delivers a better true-positive rate with a mean precision of $95.5 \pm 6.3\%$ and $96.1 \pm 3.6\%$. Using Method 2, the mean recall for MRI-A is evidently better, achieving $97.6 \pm 3.0\%$ versus $92.5 \pm 3.6\%$ using 3D-Method 2, and similarly using MRI-B achieves $99.3 \pm 0.7\%$ versus $92.9 \pm 4.2\%$.

Using CT-NIH, the approach 3D-Method 2 provides significantly better contouring compared to Method 2 by reducing the mean Hausdorff distance by approximately $2mm$ ($9.56mm$) but a higher standard deviation by approximately $2mm$ ($5.67mm$) reflects an area to improve consistency of outcome. The MRI datasets follow a similar pattern of performance. MRI-A achieves an improved mean Hausdorff distance of $10.23mm$ (a reduction by $0.32mm$ compared to Method 2) whereas MRI-B delivers a significantly better mean Hausdorff distance of $7.43mm$ (a reduction by approximately $2mm$ compared to Method 2). However, this improved mean metric can be contrasted with a higher standard deviation using 3D-Method 2 versus Method 2 as in $5.98mm$ (versus $3.72mm$) and $5.26mm$ (versus $1.98mm$) for MRI-A and MRI-B respectively.

6.5 Conclusion

In this chapter, the methods presented in Chapters 4 and 5 are evaluated and compared with one another. Method 1 and Method 2 address the intensity consistency problem that is often the case when segmenting image volumes on a slice-by-slice basis. In Method 1, a novel post-processing stage optimises tissue classification using morphological, anatomical and radiological knowledge about the connectivity between pancreatic contours. Method 2 integrates a novel loss function using the Hausdorff-distance metric to train a deep learning model that eventually predicts pancreatic features in a test image volume. 3D-Method 2 employs a 3D-based deep learning approach by combining

an organ localisation model and a segmentation model, which performs better in terms of mean quantitative accuracy but demonstrates less statistical stability than Method 2. Moving forward, the segmentation results obtained in this chapter are compared to reported state-of-the-art methods in the following chapter.

Chapter 7

Discussion and Improvements

7.1 Introduction

This chapter discusses the qualitative and quantitative pancreas segmentation accuracy scores achieved using Method 1, Method 2 and 3D-Method 2 as evaluated in Chapter 6 in comparison with reported state-of-the-art approaches. This chapter also discusses potential improvements to the proposed segmentation approaches, including developments to the deep learning architecture in Method 2 and 3D-Method 2. Section 7.2 explores the significance of the proposed methods in comparison with state-of-the-art methodologies, and proceeds to identify potential areas for optimisation, particularly concerning multiple datasets of differing modalities. Next, Section 7.3 delivers a list of optimisations in pre-processing and post-processing stages, and utilising additional deep learning blocks to focus the network on the anatomical object of interest. Section 7.4 summarises the advantages of the proposed methods, Method 1, Method 2 and 3D-Method 2 including potential improvements that could also integrate into and improve the methodology applicable to multiple organs and muscular tissue.

7.2 Discussion

With the exception of the publicly available CT dataset that has been utilised in publications [14, 30, 72, 73, 76, 77] direct comparison with other methods in literature is difficult due to differences in modality, scanner imaging protocols and spatial resolution. It should be noted that the evaluated MRI data was obtained using a scanner imaging protocol (sequence) that was not optimised for any particular organ, thus adding to the challenge of detecting the pancreas. That said, the approaches presented in Chapter 4 (Method 1) and Chapter 5 (Method 2 and 3D-Method 2) report better quantitative pancreas segmentation results in comparison to other state-of-the-art techniques such as [1, 14, 21, 23, 24, 28, 70, 133, 171, 173, 174].

Table 7.1 displays the DSC, JI, PC and RC for the CT-NIH dataset as mean \pm standard deviation [lowest, highest], in comparison to other automatic approaches reported in research literature that involve CT modality. Similarly, Table 7.2 displays the quantitative accuracy scores for datasets MRI-A and MRI-B.

7.2.1 Quantitative Assessment

Method 1: Morphological and Multi-level Geometrical Descriptor Analysis

Method 1 delivers better quantitative accuracy, comparing DSC standard deviation of 5.7% for MRI-A and 5.1% for MRI-B versus 8.7% reported in [24]. Although the mean DSC result for MRI in [30] is relatively higher, Method 1 delivers a smaller standard deviation versus 6.7% in the reported publication. By employing 3D max-flow and min-cuts approach, in combination with meticulous geometrical descriptor analysis, the proposed approach advances the quantitative performance to a mean DSC of 79.30% in testing CT-NIH and reveals statistical significance (t-test, $p < 0.0001$). This method is close to the state-of-the-arts [30, 72, 73] with approximate differences in mean DSC of 2%, 3% and 3% respectively. The mean of listed CT data standard deviations in DSC from other reported publications in Table 7.1 is 13.1%, whereas the proposed method achieves approximately one-third the value in standard deviation. In fact, the proposed method also delivers better statistical stability, comparing 4.40% versus 6.27% [72], 6.70% [30] and 5.68% [73] in the standard deviation of DSC scores. Moreover, the minimum DSC score is approximately 73% for CT-NIH, whereas [72, 73] report cases that are less than 63% in DSC and [14, 21, 25, 27, 28] report cases with DSC that are less than 10%.

Method 2: Integrating Deep Learning

Method 2 outperforms the state-of-the-art work on MRI, both in terms of accuracy in DSC and statistical stability. In comparison to Method 1, this deep learning-based approach raises the mean DSC score by 4.53% for MRI-A and 4.13% for MRI-B. Furthermore, the standard deviation drops by 0.96% for MRI-A and 2.65% for MRI-B in comparison to Method 1. The minimum DSC scores are approximately 72% for MRI-A (versus 70% in Method 1) and approximately 80% for MRI-B (versus 71% in Method 1), whereas the methods described in [24, 30] report cases with DSC that are less than 60%. The mean of listed MRI data standard deviations in DSC from other reported publications in Table 7.2 is 7.7%, whereas the proposed method achieves approximately two times lower value in standard deviation.

Method	DSC (%)	JI (%)	PC (%)	RC (%)	N
Okada, <i>et al.</i> , 2011 [33]	-	46.6	-	-	28
Ham., <i>et al.</i> , 2013 [79]	-	61.2 \pm 9.08	-	-	40
Farag, <i>et al.</i> , 2014 [78]	68.8 \pm 25.6	57.2 \pm 25.4	71.5 \pm 30.0	72.5 \pm 27.2	80
Farag, <i>et al.</i> , 2016 [1]	70.7 \pm 13.0	57.9 \pm 13.6	71.6 \pm 10.5	74.4 \pm 15.1	80
Roth, <i>et al.</i> , 2015 [14]	71.8 \pm 10.7 [25.0, 86.9]	-	-	-	82
Roth, <i>et al.</i> , 2015 [76]	68.0 \pm 10.0 [43.0, 80.0]	-	-	-	82
Roth, <i>et al.</i> , 2018 [72]	81.3 \pm 6.3	-	-	-	82
Zhou, <i>et al.</i> , 2016 [73]	82.4 \pm 5.7 [62.4, 90.9]	-	-	-	82
Cai, <i>et al.</i> , 2017 [30]	82.4 \pm 6.7 [60.0, 90.1]	70.6 \pm 9.0 [42.9, 81.9]	-	-	82
Cai, <i>et al.</i> , 2018 [130]	83.3 \pm 5.6 [59.0, 91.0]	71.8 \pm 7.7 [41.8, 83.5]	84.5 \pm 6.2 [60.7, 96.7]	82.8 \pm 8.4 [56.4, 94.6]	82
Roth, <i>et al.</i> , 2016 [77]	78.0 \pm 8.2	-	-	-	82
Xia, <i>et al.</i> , 2018 [131]	84.6 \pm 5.1	-	-	-	82
Shim., <i>et al.</i> , 2010 [32]	-	57.9	-	-	98
Wang, <i>et al.</i> , 2014 [27]	65.5 \pm 18.6 [2.4, 90.2]	-	-	-	100
Wolz, <i>et al.</i> , 2012 [29]	65.5	49.6	70.7	62.9	100
Chu, <i>et al.</i> , 2013 [25]	69.1 \pm 15.3	54.6 \pm 15.9	-	-	100
Okada, <i>et al.</i> , 2015 [28]	73.4 \pm 15.1	60.4 \pm 16.7	-	-	134
Liu, <i>et al.</i> , 2019 [134]	71.45	-	-	-	136
Saito, <i>et al.</i> , 2016 [75]	74.4 \pm 20.2	62.3 \pm 19.5	-	-	140
Oda, <i>et al.</i> , 2016 [74]	75.1 \pm 15.4	62.1 \pm 16.6	-	-	147
Kar., <i>et al.</i> , 2017 [70]	78.5 \pm 14.0	66.3 \pm 15.5	-	-	150
Tong, <i>et al.</i> , 2015 [71]	71.1 \pm 14.7	56.9 \pm 15.2	-	-	150
Wolz, <i>et al.</i> , 2013 [21]	69.6 \pm 16.7	55.50 \pm 17.1	67.9 \pm 18.2	74.1 \pm 17.1	150
CT-NIH (Method 1)	79.3 \pm 4.4 [72.8, 86.0]	66.1 \pm 6.2 [58.2, 75.5]	88.6 \pm 6.9 [74.5, 99.5]	97.6 \pm 1.9 [94.7, 99.8]	82
CT-NIH (Method 2)	83.1 \pm 5.3 [69.3, 89.6]	71.4 \pm 7.4 [53.1, 81.1]	97.6 \pm 5.3 [75.2, 99.9]	97.1 \pm 3.7 [84.9, 100.0]	82
CT-NIH (3D-Method 2)	84.7 \pm 7.9 [65.4, 95.1]	74.2 \pm 11.4 [48.7, 86.1]	87.2 \pm 5.6 [71.7, 95.3]	82.6 \pm 10.3 [60.2, 94.9]	82

Table 7.1: Overall DSC, JI, PC and RC shown as mean \pm standard deviation [lowest, highest] for automatic pancreas segmentation methods in CT modality image volumes. The value of N represents the dataset size.

Method	DSC (%)	JI (%)	PC (%)	RC (%)	N
Cai, J., <i>et al.</i> , 2017 [24]	76.1 \pm 8.7 [47.4, 87.1]	-	-	-	78
Cai, J., <i>et al.</i> , 2017 [30]	80.5 \pm 6.7 [59.1, 89.4]	67.9 \pm 8.9 [41.9, 80.9]	-	-	79
Cai, J., <i>et al.</i> , 2018 [130]	80.7 \pm 7.4 [48.8, 90.5]	68.2 \pm 9.6 [32.3, 82.7]	84.3 \pm 7.6 [55.8, 95.8]	78.3 \pm 10.2 [38.6, 95.0]	79
Asaturyan, H., <i>et al.</i> , 2018 [13]	75.5 \pm 7.0 [65.0, 86.9]	61.2 \pm 9.2 [48.1, 76.9]	-	-	130
MRI-A (Method 1)	79.6 \pm 5.7 [68.6, 87.5]	66.5 \pm 7.9 [52.4, 77.8]	85.7 \pm 11.5 [71.5, 98.9]	94.7 \pm 10.7 [50.5, 99.7]	216
MRI-A (Method 2)	84.1 \pm 4.6 [72.1, 89.6]	72.9 \pm 6.5 [56.4, 81.1]	95.5 \pm 6.3 [71.7, 99.7]	97.6 \pm 3.0 [89.9, 100.0]	216
MRI-A (3D-Method 2)	89.9 \pm 3.4 [82.4, 94.2]	81.87 \pm 5.6 [73.8, 89.1]	87.5 \pm 3.8 [78.9, 91.7]	92.5 \pm 3.6 [84.1, 9.7]	216
MRI-B (Method 1)	81.6 \pm 5.1 [71.3, 88.0]	69.2 \pm 7.1 [55.5, 78.6]	82.2 \pm 5.8 [73.1, 93.5]	84.8 \pm 8.3 [69.4, 96.0]	132
MRI-B (Method 2)	85.7 \pm 2.3 [79.9, 90.3]	75.1 \pm 3.5 [66.5, 82.4]	96.1 \pm 3.6 [86.7, 100.0]	99.3 \pm 0.7 [99.9, 100.0]	132
MRI-B (3D-Method 2)	90.2 \pm 5.1 [72.2, 95.8]	82.6 \pm 7.8 [56.4, 91.9]	88.0 \pm 6.9 [58.4, 95.7]	92.9 \pm 4.2 [81.7, 97.1]	132

Table 7.2: Overall DSC, JI, PC and RC shown as mean \pm standard deviation [lowest, highest] for automatic pancreas segmentation methods in MRI modality image volumes. The value of N represents the dataset size.

Method using CT-NIH	DSC (%)	JI (%)
Farag, A., <i>et al.</i> , 2014 [78]	68.8 ± 25.6	57.2 ± 25.4
Roth, H.R., <i>et al.</i> , 2015 [76]	68.0 ± 10.0	51.5 ± 5.1
Ronneberger, O., <i>et al.</i> , 2015: 2D UNet Model [171]	79.7 ± 7.6	66.8 ± 9.60
Long, J., <i>et al.</i> , 2015: 2D FCN Model [174]	80.3 ± 9.0	67.1 ± 4.7
Roth, H.R., <i>et al.</i> , 2016 [14]	71.8 ± 10.7	56.0 ± 5.7
Roth, H.R., <i>et al.</i> , 2016 [77]	78.0 ± 8.2	63.9 ± 4.3
Gibson, E., <i>et al.</i> , 2018 [133]	75.4 ± 9.8	60.5 ± 5.1
Zhou, Y., <i>et al.</i> , 2016 [73]	82.4 ± 5.7	70.1 ± 2.9
Cai, J., <i>et al.</i> , 2017 [30]	82.4 ± 6.7	70.6 ± 9.0
Farag, A., <i>et al.</i> , 2017 [1]	70.7 ± 13.0	57.9 ± 13.6
Xia, Y., <i>et al.</i> , 2018 [131]	84.6 ± 5.1	73.3 ± 2.6
Roth, H.R., <i>et al.</i> , 2018 [72]	81.3 ± 6.2	68.5 ± 3.2
Roth, H.R., <i>et al.</i> , 2018: 3D FCN Model [173]	76.8 ± 9.4	62.3 ± 4.9
Proposed: Method 1	79.3 ± 4.4	66.1 ± 6.2
Proposed: Method 2	83.1 ± 5.3	71.4 ± 7.4
Proposed: 3D-Method 2	84.7 ± 7.9	74.2 ± 11.4

Table 7.3: Overall DSC and JI are shown as mean \pm standard deviation (%) for automatic pancreas segmentation methods using the **publicly available CT dataset**.

Method using MRI-A	DSC (%)	JI (%)
Ronneberger, O., <i>et al.</i> , 2015: 2D UNet Model [171]	69.1 ± 10.2	53.8 ± 14.2
Long, J., <i>et al.</i> , 2015: 2D FCN Model [174]	70.2 ± 8.5	63.5 ± 13.5
Roth, H.R., <i>et al.</i> , 2016 [14]	44.5 ± 25.2	32.7 ± 29.4
Gibson, E., <i>et al.</i> , 2018 [133]	52.6 ± 17.1	44.1 ± 20.7
Roth, H.R., <i>et al.</i> , 2018: 3D FCN Model [173]	65.2 ± 10.1	52.2 ± 15.3
Proposed: Method 1	79.6 ± 5.7	66.5 ± 7.9
Proposed: Method 2	84.1 ± 4.6	72.9 ± 6.5
Proposed: 3D-Method 2	89.9 ± 3.4	81.9 ± 5.6

Table 7.4: Overall DSC and JI are shown as mean \pm standard deviation (%) for automatic pancreas segmentation methods using the **MRI-A dataset**.

Method using MRI-B	DSC (%)	JI (%)
Ronneberger, O., <i>et al.</i> , 2015: 2D UNet Model [171]	72.8 ± 7.5	67.9 ± 10.2
Long, J., <i>et al.</i> , 2015: 2D FCN Model [174]	70.9 ± 7.7	65.4 ± 13.5
Roth, H.R., <i>et al.</i> , 2016 [14]	50.1 ± 22.7	44.9 ± 12.0
Gibson, E., <i>et al.</i> , 2018 [133]	55.8 ± 18.6	49.9 ± 18.7
Roth, H.R., <i>et al.</i> , 2018: 3D FCN Model [173]	69.6 ± 11.5	61.2 ± 15.9
Proposed: Method 1	81.6 ± 5.1	69.2 ± 7.1
Proposed: Method 2	85.7 ± 2.3	75.1 ± 3.5
Proposed: 3D-Method 2	90.2 ± 5.1	82.6 ± 7.8

Table 7.5: Overall DSC and JI are shown as mean \pm standard deviation (%) for automatic pancreas segmentation methods using the **MRI-B dataset**.

Applying Image Data to State-of-the-Art Segmentation Models

The proposed methods are also evaluated against the direct performance of multiple baseline segmentation models, which have served as the foundation for many other state-of-the-art segmentation approaches. Table 7.3 displays the quantitative segmentation results (DSC and JI) using the CT-NIH dataset against other (mostly) deep learning based methods. Similarly, Table 7.4 and Table 7.5 display the quantitative segmentation results (DSC and JI) using MRI-A and MRI-B, respectively. As highlighted in bold across all tables, Method 2 produces mean quantitative accuracy scores that outperform or are close to the state-of-the-art. Furthermore, Method 1 and Method 2 produce a relatively lower standard deviation in comparison to other methods, reflecting stability in the proposed techniques.

One of the drawbacks of U-Net [171] is that network optimisation may be delayed at the stage containing the middle layers, posing a risk of ‘ignoring’ layers with abstract features during network optimisation. Furthermore, while several layers of max-pooling in U-Net achieves more translation invariance, there is also a loss of spatial resolution in the feature maps, posing a threat for accurate segmentation where pancreas boundary delineation is essential. In contrast, the architecture of Method 2 ensures that boundary information is stored in the encoder feature maps before any sub-sampling, retaining only the max-pooling indices. When the decoder network upsamples its input feature maps using the memorised max-pooling indices from the corresponding encoder feature maps, sparse feature maps are produced and then convolved with a decoder filter to generate dense, deep feature maps containing improved contextual information about the pancreas’ boundary.

The method in [14] lacks the advantage of encoder-decoder architecture. Furthermore, while the FCN [174] upsamples only once with one layer in the decoder, the encoder-decoder architecture in Method 2 has multiple upsampling layers and employs skip connections and concatenates instead of adding. The FCN suffers from the drawback of not capturing local information such as the pancreas shape and structural differences. Moreover, although FCNs offer the advantage of producing detailed feature maps by capturing pancreatic spatial contextual information, there is the subsequent risk of network overfitting, often extremely sensitive to alternations in the pancreas' spatial context and structure. Also, employing the Cross-entropy loss function using FCN has a drawback on predicting true-positive pixels, while the proposed Hausdorff-Sine loss achieves a better prediction of true-positive pixels and delivers strong discrimination of true-negative pixels.

The loss of image quality due to the extreme downsampling of input image volumes to spatial sizes $144 \times 144 \times 144$ [133] and $132 \times 132 \times 116$ [173] impact the final segmentation outcome. Moreover, the Dice loss in [133] squares the convolutional weights in the backpropagation stage, causing instability during network optimisation and a higher rate of false-negative predictions. In contrast, the Hausdorff-Sine minimises the maximum deviation between a prediction and desired outcome; and the addition of a sinusoidal component, inspired by the Cross-entropy loss, increases non-linearity to reduce over-fitting during network training.

As an extension of the second proposed approach, 3D-Method 2 comparatively outperforms state-of-the-art approaches in CT-NIH and is relatively comparable to [131] albeit with a higher standard deviation. This 3D-based deep learning approach surpasses both Method 1 and Method 2 in terms of DSC in MRI. With that said, the evaluation of MRI-B presents high statistical variation among distinct MRI test cases (volumes), forcing a poorer mean precision and recall accuracy score.

7.2.2 Qualitative Assessment

In Method 1, the primary source of segmentation error relates to the remaining surrounding pancreas tissue, including the superior mesenteric vein, splenic artery and duodenum. In contrast, some test image volumes have borderline, pancreas boundary under-segmentation using Method 2. Qualitative feedback from an expert radiographer and radiologist confirms that Method 2 produces detailed pancreas contouring for every protrusion and indentation as opposed to an approximate or mean tracing of the organ. Please refer to Appendix C for further information.

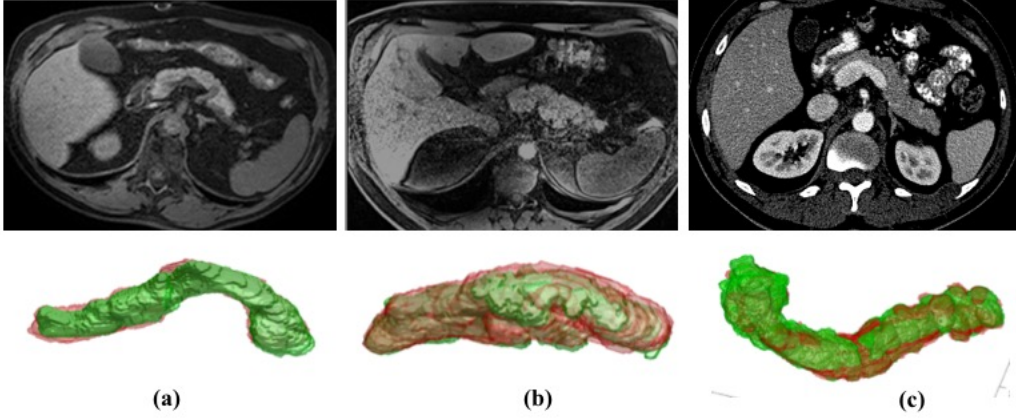


Figure 7.1: Columns (a) (b) and (c) highlights sample slice and corresponding segmentation outcome from MRI-A, MRI-B and CT-NIH respectively. Notice the variation in noise and image distortion between (a) and (b), and the difference in blur and sharpness between (a) and (c). All segmentation results (green) are mapped over the ground-truth (red) and have approximately 85% accuracy in DSC.

Robustness to Noise Variation

Unlike other publications that focus on one modality or one dataset per modality, both proposed methods are robust to variations in noise, distortion, sharpness and changes in greyscale intensity distribution within multiple MRI datasets and a CT dataset. Figure 7.1 highlights three different slices from exemplar test volumes and their corresponding segmentation result for Method 1. Notice the difference in levels of original noise and the comparable, accurate segmentation outcome of approximately 85% in DSC. Since Method 2 also employs a similar stage of digital contrast enhancement and continuous max-flow and min-cuts segmentation in a 3D domain, this approach is equally robust to image quality variation.

7.2.3 Computational Efficiency

Using Method 1, the overall mean runtime per CT and MRI volume is 30 and 22 minutes respectively, and not 2 to 4 hours as in [21] via an eight Intel Xeon cores clocked at 3 GHz and 32 GB RAM, [25] and [70] via an Intel Xeon Dual or Quad Core 1.86–3.07 GHz CPU. Except for training a random forest for just under 10 minutes to identify the major pancreas region, the approach described in Chapter 4 evaluates an image volume in 10 to 35 minutes via a workstation with i7-5930k-CPU at 3.50 Ghz.

In contrast to Method 1, although Method 2 involves training a network for approximately 11 hours, the testing aspect is reduced to 5 - 10 minutes. The training time for 3D-Method 2 (which employs 3D vectors) is approximately 1.5 times that of Method 2 (which employs 2D vectors). The testing time, however, ranges between 30 seconds and 1 minute, which is a significant advantage in comparison to Method 2. Furthermore,

the proposed approaches do not involve the computationally expensive costs of training a network in approximately 2 - 3 days [14, 173], 1 day [171, 174] and 21 hours [133], before segmenting a test image volume.

Therefore, while Method 1 removes the computational costs of training a network, Method 2 and 3D-Method 2 reduce the testing time and provide higher quantitative accuracy in segmentation results. In the future, these runtimes can be potentially reduced by a factor of 10 by using one or more GeForce Titan X GPUs.

7.3 Potential Optimisation

Throughout the development of the proposed approaches, the qualitative feedback has focused on boundary delineation and identification of the major region of interest in a scan. For example, organ segmentation methods that minimise the Hausdorff or average surface distance are appropriate for extended computation of organ volume and curvature. However, alternative qualitative feedback has focused on the major region of interest to highlight the “problem-area” for enhanced analysis. Therefore, novel computational techniques can be tailored to meet variable criteria that are often dependent on the primary objective of a biomedical research-based task with a clinical translation. Using Method 2 or 3D-Method 2 as the baseline reference, the following modifications can be taken into consideration to improve the boundary delineation, and reduce false-positive and false-negative “pancreas” results:

- Removal of surrounding splenic artery and vessels *prior* to the max-flow and min-cuts segmentation stage. As reported in [124], a number of clinical support structures are implemented to detect the splenic artery by optimising the spatial relationship between the vessel path, and local pancreas texture and position. Although the modality of choice is CT and thus gives a relatively more apparent distinction between organs and vessels in comparison to MRI, the concept can still extend to MRI data. Therefore, incorporating further anatomical knowledge relating to abdominal organs, in combination with radiological information about the pancreas and surrounding tissue, can potentially raise the overall accuracy score.
- Incorporate texture descriptors into a statistical model or a deep learning network that can identify key duodenum landmarks in an image *prior* to pancreas segmentation. Therefore, such a pre-processing model that is generated using heavy data augmentation can potentially minimise false-positive predictions, and even reduce computation time at the central segmentation and post-processing stage. In some test image volumes, the duodenum, which touches the head of the pancreas, often ‘sticks’ to the resultant segmentation; this occurs whenever a distinction cannot

be made between the greyscale intensities of the pancreas and duodenum at the max-flow and min-cuts segmentation stage. Often, the subtle detail separating the pancreas head and duodenum are detected by medical experts through analysis of the image slices in a back-and-forth manner, whereby identifying the features or slight contours required to establish organ separation. In fact, it can be possible to ‘strengthen’ the detection of weak boundaries between the duodenum and pancreas head in a manner that is similar to the approach in [129] for the renal cortex and kidneys. Here, a statistical shape model for the area of interest is developed in order to initialise the outer surface and inner surface automatically. Afterwards, a graph-cuts algorithm can be introduced to detect the optimal boundary of the organ.

- It is also the intention to train a deep learning model combining multiple modalities, including a CT dataset and multiple MRI datasets, each of which relate to a different MRI sequence. The advantages of producing one deep learning model that provides relatively accurate predictions are cost-effective and also, on a developmental level, can allow the programmer to observe the features that the deep learning model is learning across multiple modalities and MRI sequences.

7.3.1 Improvements to Architecture in Method 2

It is noted that the proposed “targeted-to-fine” segmentation approach in Method 2 can be extended to *enhanced* targeted location detection as in [72, 73], and therefore improved segmentation performance may be achievable with a combination of these methodologies. This is particularly important for detecting smaller pancreas features in 2D (slice).

One of the motivations for modifying the (SegNet) deep learning model is to reduce the memory and runtime in the upsampling (decoder) stage while maintaining targeted, quality pancreas segmentation. Such requirements are driven by the need to deliver accurate and efficient segmentation between the pancreas and surrounding non-pancreas tissue. To further optimise the deep learning model, one possible strategy reduce the number of filters in the last two stages of down-sampling (encoder) from 512 to 256. Such a reduction may achieve a significant drop in classification accuracy. However, the deep learning model should ideally borderline over-segment the main pancreas region, after which the continuous max-flow and min-cuts stage ‘carves’ out the finer, more detailed pancreas boundaries. Increasing the current data augmentation to include cropping, spatial augmentation and colour augmentation might increase the accuracy classification in terms of raising the level of true-positive “pancreas” for relatively smaller pancreatic features on a slice-by-slice basis. Thus, reducing the number of filters and

increasing the amount of data augmentation might counter-balance the overall network training time.

It should be noted that one of the disadvantages of ReLU is not being differentiable at 0. Furthermore, this activation function can, at times, cause “dead neurons” during training, where the weights in the filters are not updated to different values. In order to address and overcome these challenges, activation functions such as Leaky ReLU [218], Exponential Linear Unit (ELU) [219] and Scaled Exponential Linear Unit (SELU) [220] will be investigated. Furthermore, instead of randomly initialising the weights of filters in the convolutional layers, it might be a useful practice to employ custom weight initialisation techniques [220, 221, 222] that take into account the training dataset size and the nature of the training data’s spatial resolution.

7.4 Conclusion

The comparison to state-of-the-art methodologies reveals significant advantages of the methods proposed in this thesis, both in terms of quantitative accuracy and qualitative feedback. The final quantitative pancreas segmentation results (especially using Method 2 and 3D-Method 2) are better than or close to state-of-the-art approaches for both CT and MRI modality, and report higher statistical stability with lower standard deviation metrics. The feedback obtained by an expert radiographer and an expert radiologist, reveals that the proposed methods produce detailed contouring of the pancreas for every protrusion and indentation. Given the wide variation in datasets, Method 2 or 3D-Method 2 can be further optimised and applicable to other abdominal MRI and CT image volumes and also, generalisable to other organ or muscular tissue segmentation tasks.

Chapter 8

Generalisation and Future Work

8.1 Introduction

In the past few decades, publications in medical and biomedical science have highlighted varying correlations between morphological features of organs and tumours, including volume, curvature and associated medical conditions. However, limitations to such studies include a small patient dataset, e.g. more commonly multiples of 10 patient cases, as opposed to multiples of 100 or 1000 that could otherwise generate a more reliable, accurate correlation between morphological features that can serve as a prognostic biomarker for different medical conditions.

This chapter explores the generalisation of the proposed segmentation methodologies (Method 2 and 3D-Method 2), with application to other abdominal organs and muscular tissue. Furthermore, this chapter explores potential future work of applying the proposed segmentation approaches to a broader spectrum of modalities, MRI sequences and organs. Generalisability is a significant step forward in medical image computing and biomedical research; this can serve as a tool that potentially identifies key prognostic biomarkers about a medical condition, and reduce the need for manually segmenting 3D radiological scans. A proposed generalisation framework, that is proposed in this chapter, is fully automatic and employs image volumes obtained from multiple MRI sequences. Section 8.2 addresses the biomedical and potential healthcare impact of extracting morphological features in 3D radiological scans, and also the diverse state-of-the-art publications that draw correlations and prognostic biomarkers from these morphological features in different organs and muscular tissue. Section 8.3 presents and discusses the advantages of a generalised segmentation framework that incorporates the automatic segmentation methodology discussed in Chapter 5. As an extended proof of generalisability, Sections 8.4, 8.5 and 8.6 present and discuss the experimental results for the liver, iliopsoas muscles and pediatric kidney segmentation in diverse MRI sequences. Section 8.7 proceeds to discuss the extension of the proposed segmentation

approaches in different MRI sequences, and other organ abnormalities in which combining different MRI sequences may improve the accuracy of detecting (and classifying) the abnormality of interest. Section 8.8 provides a conclusion for this chapter.

8.2 Impact of Extracting Morphological Features

Extracting accurate morphological features in organs can generate the following positive contributions to scientific knowledge, clinical research and therapy planning:

- (a) Assess and correlate changes in organ shape, volume and curvature. Please refer to Appendix D for further information.
- (b) Foster clinical studies: establish possible correlations between factors such as organ volume, organ curvature, anthropometric measures and health status¹.
- (c) Understand the relationship between organ size, structure and cancerous tumours (e.g. ductal adenocarcinoma).
- (d) Establish severity or progression of a condition (e.g. type 2 diabetes) or identification of neoplasms (tumours).
- (e) Stratification of patients in order to develop targeted treatment plan.

Having discussed the advantages and disadvantages of differing modalities in Chapter 2 (Section 2.3) it is an important reminder that MRI offers significant advantages concerning soft tissue imaging optimisation. The report published in [110] describes the importance of using MRI imaging for contouring or delineating a tumour within the pancreas, especially as it relates to cancer treatment involving 3D conformal radiation planning. Abdominal MRI sequences are complex and vary, with dependence on the scanner imaging protocol and the organ of interest, which can be very challenging for the radiation oncologist. Therefore, an automatic segmentation tool that produces accurate (detailed) organ and tumour contouring can remove the human-error variation that often occurs in manual segmentation. Moreover, a computing-based approach for automatic segmentation can offer a potentially universal standard in defining an organ or tumour.

Over the past few decades, multiple studies have highlighted the correlation between liver volume and shape, and higher body weight and alcohol consumption [111, 112]. In fact, [113] reports significant changes in liver and spleen volumes in alcoholic liver disease. Moreover, the liver volume has served as a useful measure when determining

¹Health status refers to a person's level of physical health and illness, and their sense of well-being while factoring in the possible presence of biological or physiological impairment and their history of medical conditions.

liver transplantation in both children and adults [114]. Liver volume may also correlate with several metabolic processes, particularly in relation to liver function for patients with chronic hepatitis B and C [115].

In the past few years, publications in biomedicine have shown a correlation between kidney volume and polycystic kidney disease, particularly concerning increases in cyst volume and change in glomerular filtration rate² [116, 117]. Further studies highlighted that total kidney volume (TKV) is related to complications of Autosomal Dominant Polycystic Kidney Disease (ADPKD) as well [117], and that height-adjusted TKV is a prognostic biomarker in ADPKD [116].

Information about the texture of the spleen can be used as a predictor for stages of Sickle Cell Disease³ [118] and parasitic infections [119]. The spleen is highly variable and can have a wide range of anomalies in relation to shape, size and location [120]. Although such anomalies are present from birth (such as lobulations, notches, and clefts), there are also acquired types: Splenosis⁴ is caused by trauma and an enlarged spleen can be a symptom of Sickle Cell Disease. Furthermore, [121] reported that an illness characterised by polyarthrititis, fevers, and sweats was associated with spleen enlargement and fluctuations in size.

In other more recent studies, data about the structure and volume of psoas muscles, which forms a major composite part of the iliopsoas muscles, can support subject stratification and serve as a predictor of outcome for patients treated for bladder cancer chemotherapy[122] and ovarian cancer [123].

8.3 Proposed Generalised Segmentation Framework

The proposed segmentation technique could be extended to other organs or muscles of high variability. A generalisation of the segmentation framework can serve as a tool for classification of measures, and therefore provide an indication or prognostic biomarker about a condition, or progression or severity of a condition. Such classifications would be, nonetheless, founded on supporting biomedical statistical analysis. Figure 8.1 highlights the proposed generalised segmentation framework, starting with a medical image scan as an input, with the final output as a likely predictor of a condition, disorder or disease.

²Glomerular filtration rate (GFR) is a test used to check how well the kidneys are working.

³Sickle Cell Disease is the name for a group of inherited conditions that affect the red blood cells.

⁴Splenosis is the result of spleen tissue breaking off the main organ and implanting at another site inside the body.

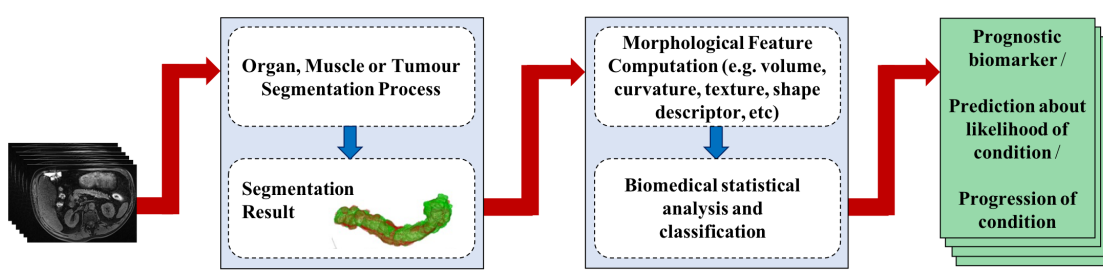


Figure 8.1: Overview of generalised segmentation framework: the initial input is a 3D medical image scan that is processed to segment an organ, muscle or tumour. The segmentation result is processed through a biomedical statistical and classification analysis model, which consequently outputs the final prediction to indicate the likelihood of a condition or progression of a current condition.

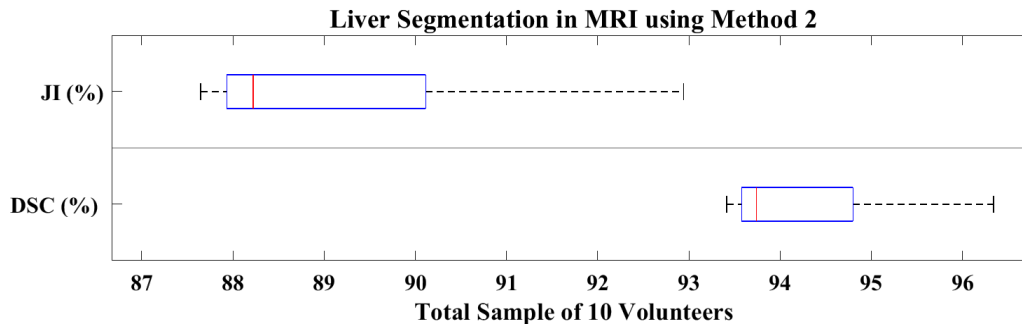


Figure 8.2: Dice similarity coefficient (DSC) and Jaccard index (JI) box plots for liver segmentation in MRI using Method 2.

8.3.1 Dataset

The dataset used in the experiments described in Sections 8.4 and 8.5 contain 30 T2-weighted abdominal 3D MRI scans (image volumes). Every MRI scan has been obtained using a Siemens Trio 3T scanner. These scans were taken from volunteers aged 18 or over, who displayed early signs of type 2 diabetes. Every 3D volume consists of 370 axial slices, with each slice of spatial size 224×174 pixels with $3mm$ thickness, and $2.2321mm$ in voxel width and height. In these experiments, the training and testing dataset is split into 20 and 10 image volumes, respectively with a 3-fold cross-validation.

8.4 Liver

This section details the analysis of the liver in MRI using Method 2 and 3D-Method 2.

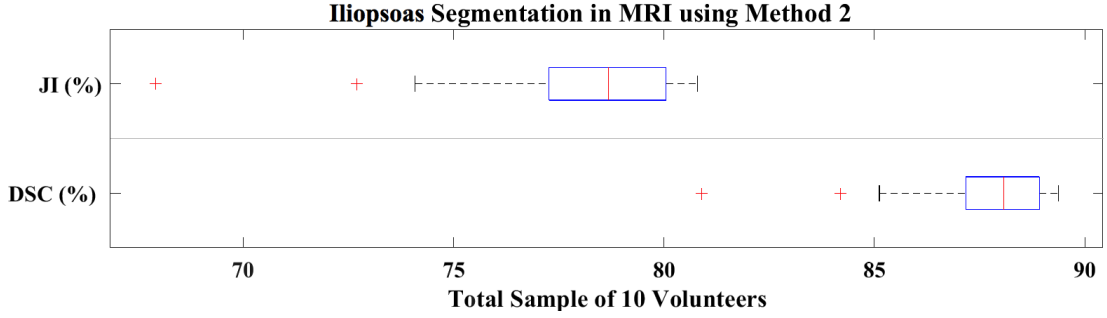


Figure 8.3: Dice similarity coefficient (DSC) and Jaccard index (JI) box plots for iliopsoas muscles segmentation in MRI using Method 2.

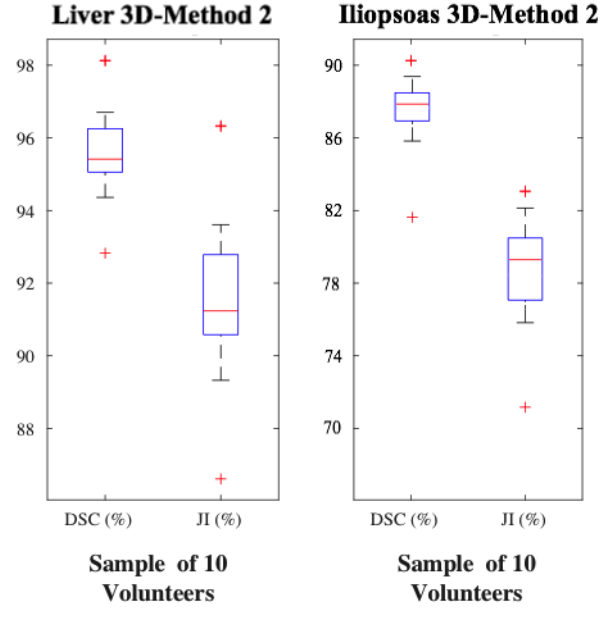


Figure 8.4: Dice similarity coefficient (DSC) and Jaccard index (JI) box plots for liver and iliopsoas muscles segmentation in MRI using 3D-Method 2.

8.4.1 Experimental Results and Analysis

A number of state-of-the-art approaches have achieved mean Dice similarity coefficient (DSC) scores of 94% [168], 94.5% [169], 89% [170], 72.9% [171] and 93.73% [172]. The methodology proposed in Method 2 both surpasses and achieves close to state-of-the-arts with a mean DSC of $94.32 \pm 0.97\%$ and a mean Jaccard index (JI) of $89.26 \pm 1.75\%$, and the low standard deviation in DSC implies that the approach is statistically robust. Figure 8.5 displays the visualisations of six different liver segmentation results in the MRI dataset, and Figure 8.2 displays the corresponding box and whisker plot representation of the segmentation results. A mean precision of $94.81 \pm 3.05\%$ and a mean recall of $97.10 \pm 1.97\%$ reflects a high true-positive rate and a low false-negative rate. Furthermore, the mean HSD is $9.24 \pm 2.84mm$, which is relatively low and better in comparison to the majority of reported organ or tumour HSD results ($> 10mm$).

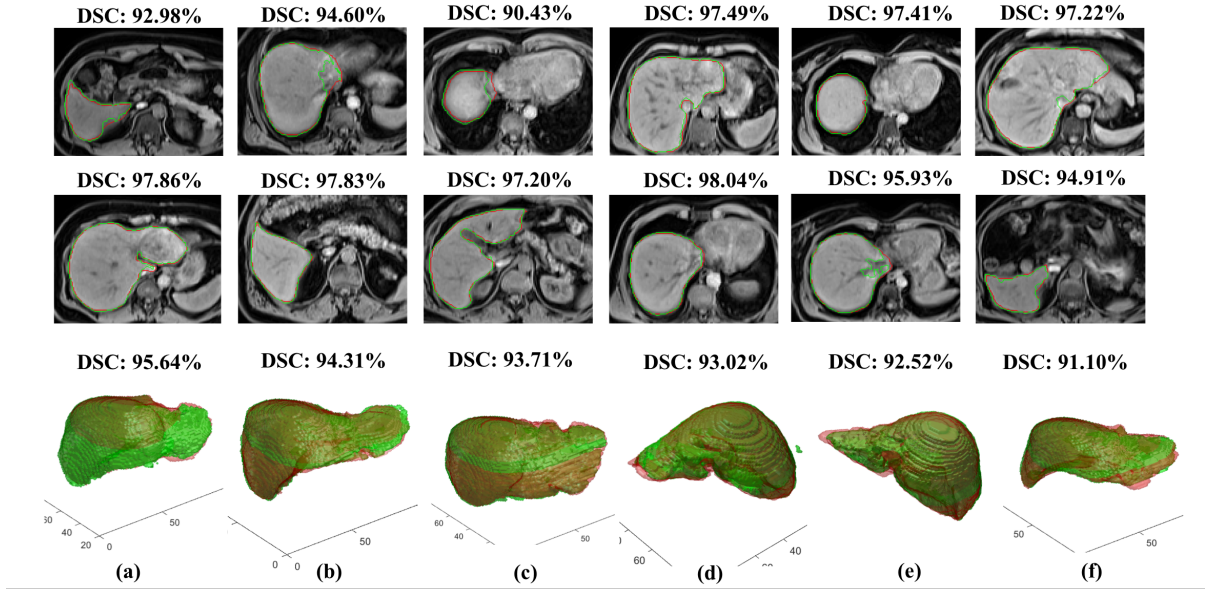


Figure 8.5: Segmentation results in six different MRI image scans (volumes) using Method 2. Every column corresponds to a single MRI volume. From left, first two rows display sample MRI axial slices with segmentation outcome (green) against ground-truth (red), and computed DSC; third row displays 3D reconstruction of entire liver (green) segmentation against its ground-truth (red) with computed DSC.

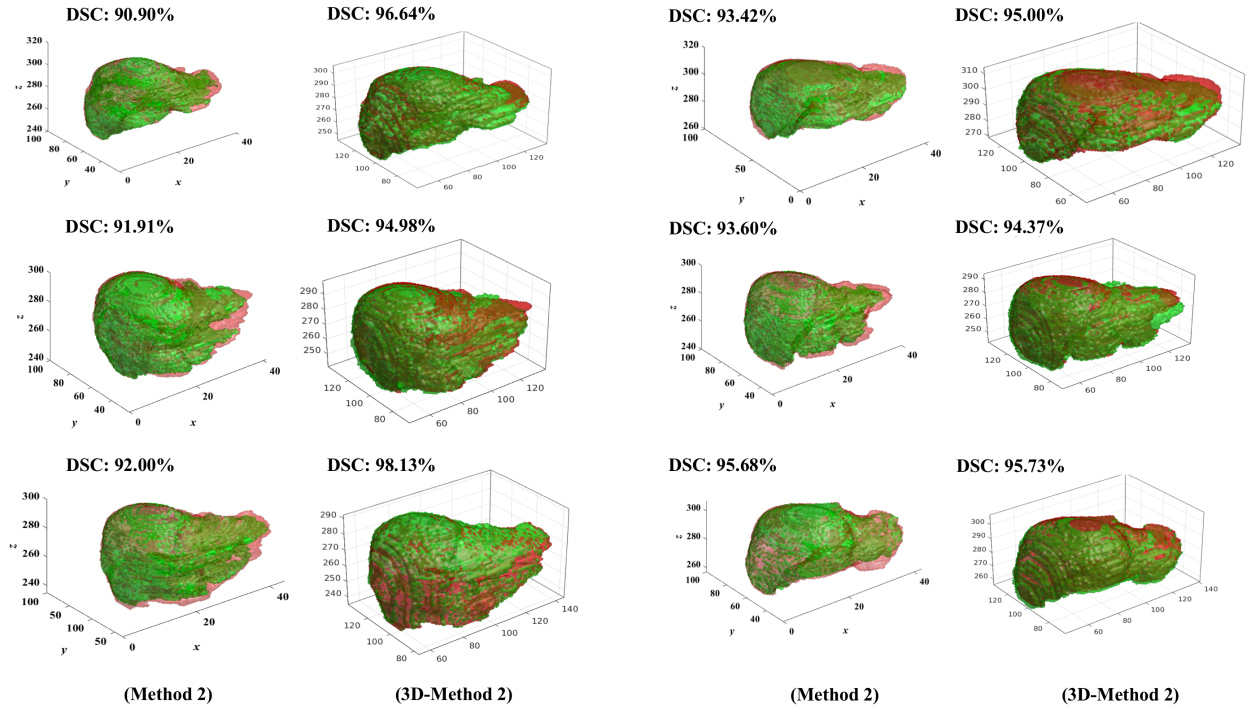


Figure 8.6: Liver organ segmentation results in six different MRI image scans (volumes) using Method 2 and 3D-Method 2. Every row of two 3D liver reconstructions highlights the segmentation results and DSC in the same MRI volume using Method 2 and 3D-Method 2 respectively. The segmentation outcome (green) overlaps the ground-truth (red) with computed DSC.

In contrast to Method 2, the proposed 3D deep learning approach (3D-Method 2) achieves a mean DSC of $95.64 \pm 1.31\%$ and mean JI of $91.66 \pm 2.42\%$, with this simultaneous rise in mean DSC (1.32%) and standard deviation (0.34%) reflecting a capability to better address certain diversity of artefact-prone imaging. Figure 8.6 displays the visualisations of six different liver segmentation results in the MRI dataset against the results obtained using Method 2, and Figure 8.4 displays the corresponding box and whisker plot representation of the segmentation results. Furthermore, the mean HSD is $5.64 \pm 2.48mm$, a promising reduction in error of approximately $4mm$ relative to Method 2. A mean precision of $94.73 \pm 2.34\%$ is approximately 1% less in standard deviation compared to Method 2, and a mean recall of $96.60 \pm 1.18\%$ reflects a maintained high true-positive rate and a low false-negative rate.

8.5 Iliopsoas Muscles

The iliopsoas muscles are a pair of muscles, each referring to the joining of the psoas (major) muscle and iliacus muscle located toward the inner hip. This section details analysis of iliopsoas muscles in MRI using Method 2 and 3D-Method 2.

8.5.1 Experimental Results and Analysis

The approach reported in [142] delivers psoas muscles segmentation results of 72.3% in mean Jaccard index (JI). In contrast, the approach proposed in Method 2 raises the state-of-the-art segmentation results for the closely compared anatomical structures to $87.58 \pm 2.05\%$ in mean DSC and $77.96 \pm 3.15\%$ in mean JI. Figure 8.7 displays the visualisations of six different iliopsoas muscles segmentation results in the MRI dataset. Very few outliers are highlighted in Figure 8.3, which displays the corresponding box and whisker plot. The mean HSD is $5.42 \pm 1.18mm$, which, similar to the liver, is relatively low and better in comparison to the majority of reported organ or tumour HSD results. Furthermore, the mean precision result is an impressive $99.32 \pm 0.52\%$ and the mean recall is $96.85 \pm 1.93\%$ reflecting a high true-positive rate and low false-negative rate.

3D-Method 2 delivers a relatively better performance in mean DSC as in $88.41 \pm 2.39\%$ and a mean JI of $78.56 \pm 3.59\%$. Figure 8.8 displays the visualisations of four different iliopsoas muscles segmentation results in the MRI dataset against the results obtained using Method 2, and Figure 8.4 displays the corresponding box and whisker plot representation of the segmentation results. Arguably, the liver and iliopsoas muscles segmentation results using the 3D deep learning approach are not significantly better than the 2D based approach in Method 2; that said, it is necessary to note the advantages of downsampling the image volumes for computational efficiency and to optimise for classification between clinically healthy and abnormal morphological features.

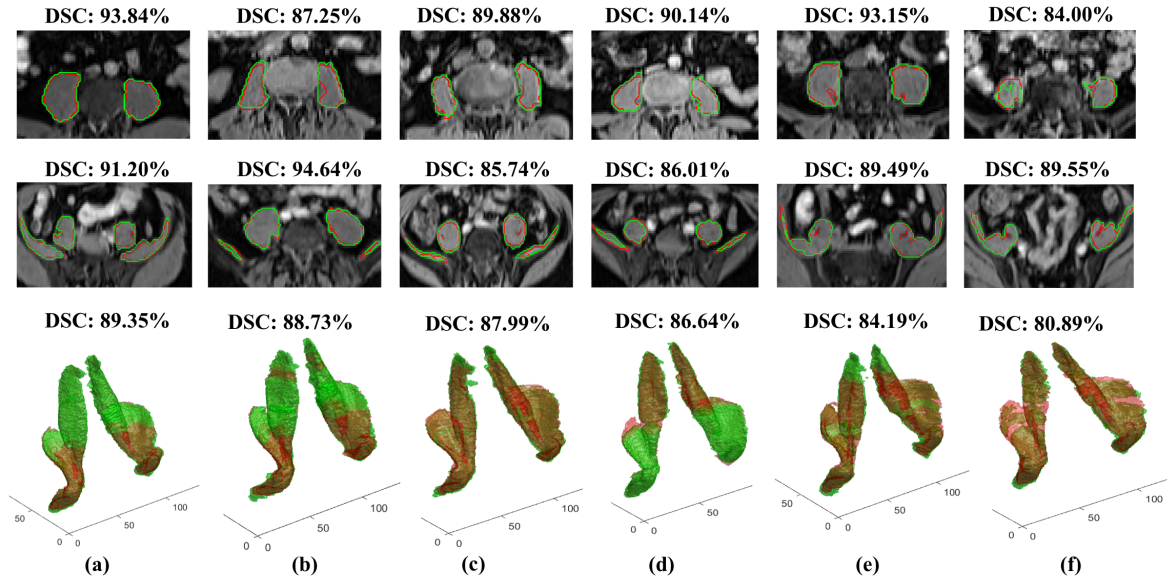


Figure 8.7: Segmentation results in six different MRI image scans (volumes) using Method 2. Every column corresponds to a single MRI volume. From left, first two rows display sample MRI axial slices with segmentation outcome (green) against ground-truth (red) and computed DSC; third row displays 3D reconstruction of the entire iliopsoas muscles (green) segmentation against its ground-truth (red) with DSC.

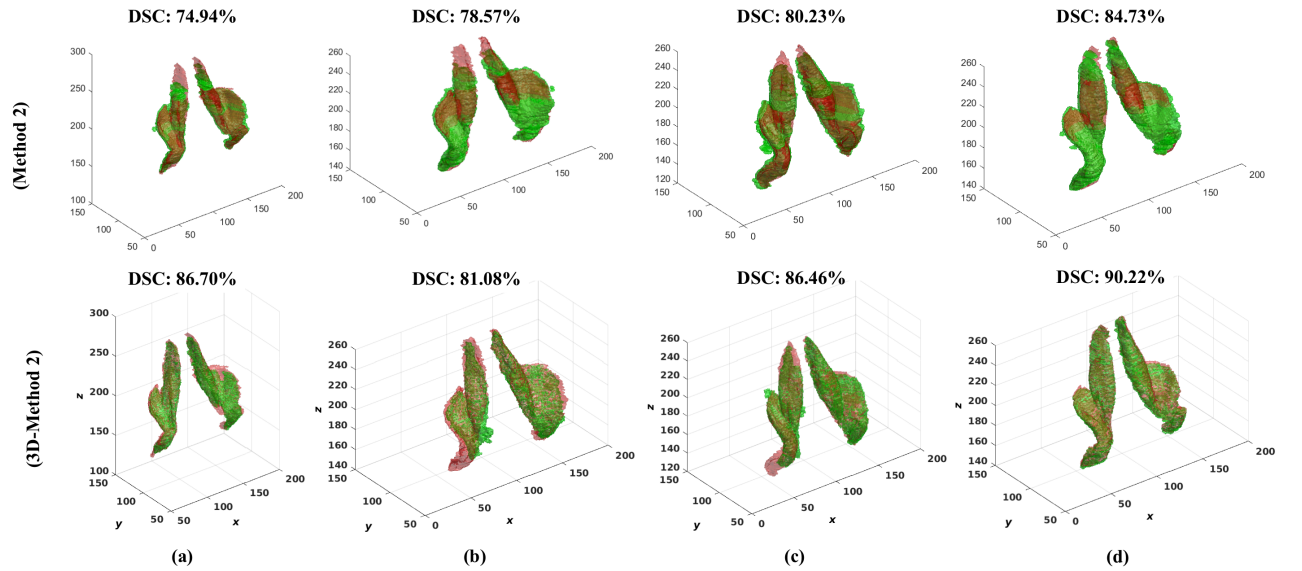


Figure 8.8: Iliopsoas muscles segmentation results in four different MRI image scans (volumes) using Method 2 and 3D-Method 2. Every column of two 3D iliopsoas muscles reconstructions highlights the segmentation results and DSC in the same MRI volume using Method 2 (top row) and 3D-Method 2 (bottom row). The segmentation outcome (green) overlaps the ground-truth (red).

8.6 Kidneys

This section covers the segmentation results and analysis of clinically “normal” and “abnormal” pediatric kidneys in Dynamic Contrast Enhanced (DCE) MRI using 3D-Method 2.

8.6.1 Dataset

Thirty-two 4D DCE-MRI scans of pediatric patients were acquired at 3 Tesla (T) for 6 minutes after injecting Gadavist⁵ using a motion-robust, radial stack-of-stars 3D FLASH sequence as follows: Repetition Time (TR)⁶ of 3.56 milliseconds (*ms*), Echo Time (TE)⁷ of 1.39 *ms* and Flip Angle (FA)⁸ of 12 degrees. A 4D volume can be viewed as a stack of 3D volumes of the same region of interest captured over a period of time, and therefore a 4D scan has a temporal dimension. Every 3D volume consists of 32 coronal slices, with each slice of spatial size 224×224 pixels and 3*mm* thickness, and 1.25*mm* in voxel width and height. The dataset includes abdominal scans from a broad spectrum of ages (2 months through to 17 years) with varying kidney conditions. Sixteen scans were taken from patients with hydronephrosis⁹, who had received MRI scanning as part of their clinical protocol since 2017. This particular group depicted kidney anatomy that deviates significantly from a clinically healthy shape due to varying levels of hydronephrosis. The remaining number of scans were taken from patients recruited under a protocol approved by the Institutional Review Board¹⁰. This group of patients had already undergone contrast-enhanced MRI clinically, but they also received 6 additional minutes of research-based imaging of their kidneys within the same session to acquire the DCE-MRI. The acquisition protocol was optimised to achieve a mean temporal resolution of 3.3 seconds for the arterial phase¹¹ (2 minutes) and 13 seconds for the remaining phase (4 minutes). The 4D dynamic imaged series were reconstructed offline using a compressed sensing algorithm to subsequently reduce streaking artefacts [177].

⁵Gadavist is a gadolinium-based MRI contrast agent.

⁶In MRI scanning, the Repetition Time (TR) is the time between consecutive radio frequency (RF) pulses applied to the same 2D slice.

⁷In MRI scanning, the Echo Time (TE) is the time between the delivery of the radio frequency (RF) pulse and the peak of the signal induced in the MRI machine’s coil.

⁸In MRI scanning, the Flip angle (FA) refers to the degree of rotation the net magnetisation vector exhibits during the delivery of a radio frequency (RF) pulse.

⁹Hydronephrosis is a condition that normally occurs when a kidney swells because urine fails to sufficiently drain from the kidney to the bladder.

¹⁰The Institutional Review Board (IRB) is a committee recognised to protect the rights and safety of human research subjects who participate in research activities at an institution, and to ensure that methods proposed for research at the affiliated institution are ethical.

¹¹In a medical scanning acquisition, the arterial phase refers to a time range after contrast injection, in which the acquired image shows full enhancement of hepatic arteries and lacks the antegrade (forward-moving) enhancement of hepatic veins.

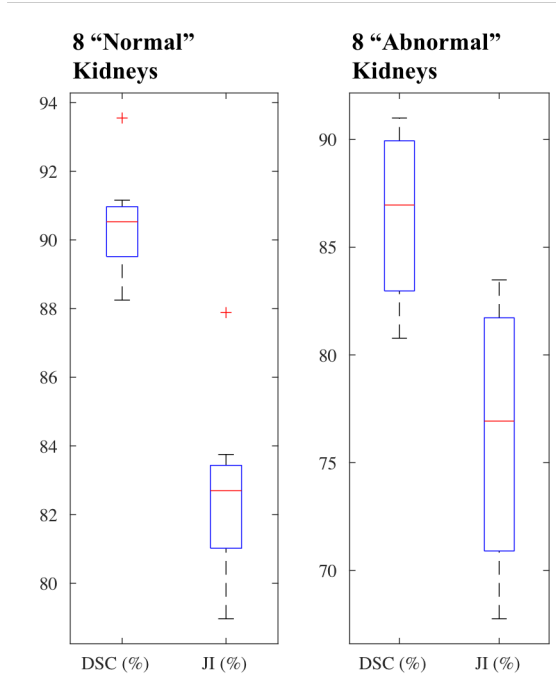


Figure 8.9: Dice Score coefficient (DSC) and Jaccard index (JI) box plots for "abnormal" and "normal" kidneys segmentation in MRI using 3D-Method 2.

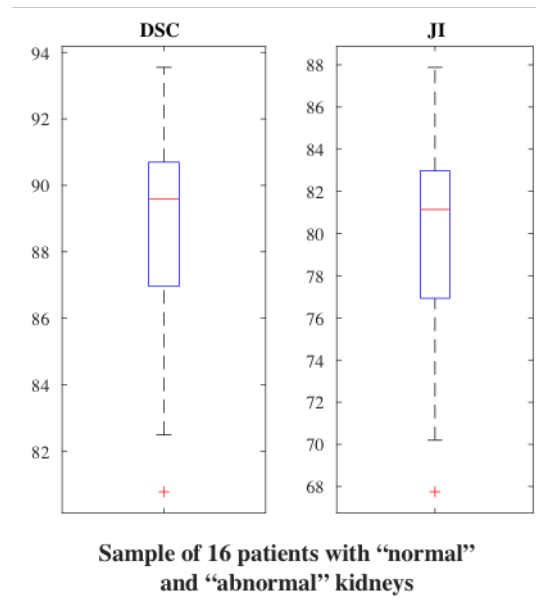


Figure 8.10: Dice Score coefficient (DSC) and Jaccard index (JI) box plots for "abnormal" and "normal" kidneys segmentation combined in MRI using 3D-Method 2.

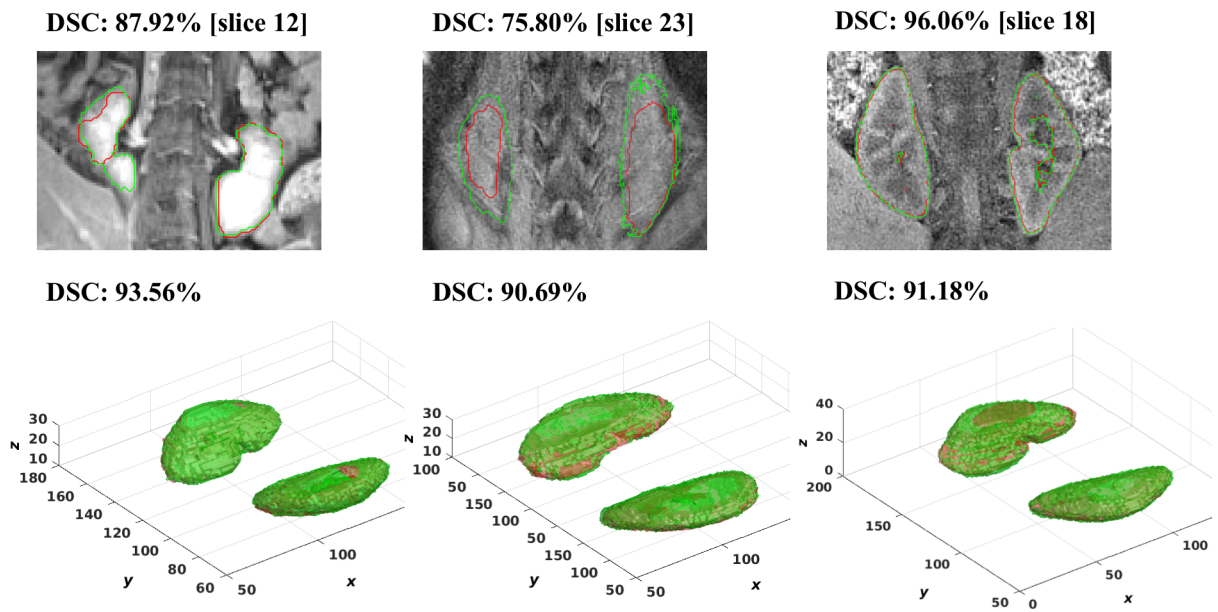


Figure 8.11: Using 3D-Method 2, a first sample batch of segmentation results in three different DCE-MRI scans (4D volumes) depicting clinically “normal” kidneys. Every column corresponds to a single DCE-MRI 4D volume. Top row displays sample slices with segmentation outcome (green) that overlap the ground-truth (red) and DSC; bottom row displays 3D reconstruction of whole kidneys segmentation that overlaps the ground-truth and DSC.

8.6.2 Experimental Results and Analysis

In the following experiments, the training and testing dataset is split into 24 and 8 image volumes, respectively. The training dataset combines scans depicting both “normal” and “abnormal” kidneys. There are two test datasets, the first of which contains 8 DCE-MRI scans that show healthy kidneys; and the second test dataset contains 8 DCE-MRI scans depicting diverse abnormalities that result in irregular shape, size and location. It should be noted that two different networks are trained, with each network omitting the corresponding test data.

Figures 8.11 and 8.12 show altogether six different 3D reconstructions for “normal” kidneys, in addition to an accompanying slice that highlights the boundary contouring. Figures 8.13 and 8.14 provide similar information for six different “abnormal” kidneys cases. Using 3D-Method 2 and evaluating the “normal” dataset achieves a mean DSC of $90.48 \pm 1.56\%$ and a mean JI of $82.64 \pm 2.61\%$. The relatively low standard deviation indicates robustness towards intensity variation of contextual information. A relatively high mean precision of $86.62 \pm 2.26\%$ and a mean recall of $94.74 \pm 1.98\%$ indicates that 3D-Method 2 minimises the false-negative prediction rate. Despite the high variation in the age of the patients who underwent the scanning - and thus the shape of the kidneys - analysing the boundary delineation accuracy achieves a mean HSD of $17.43 \pm 4.79mm$. In contrast to the performance of 3D-Method 2, the 3D U-Net performs significantly

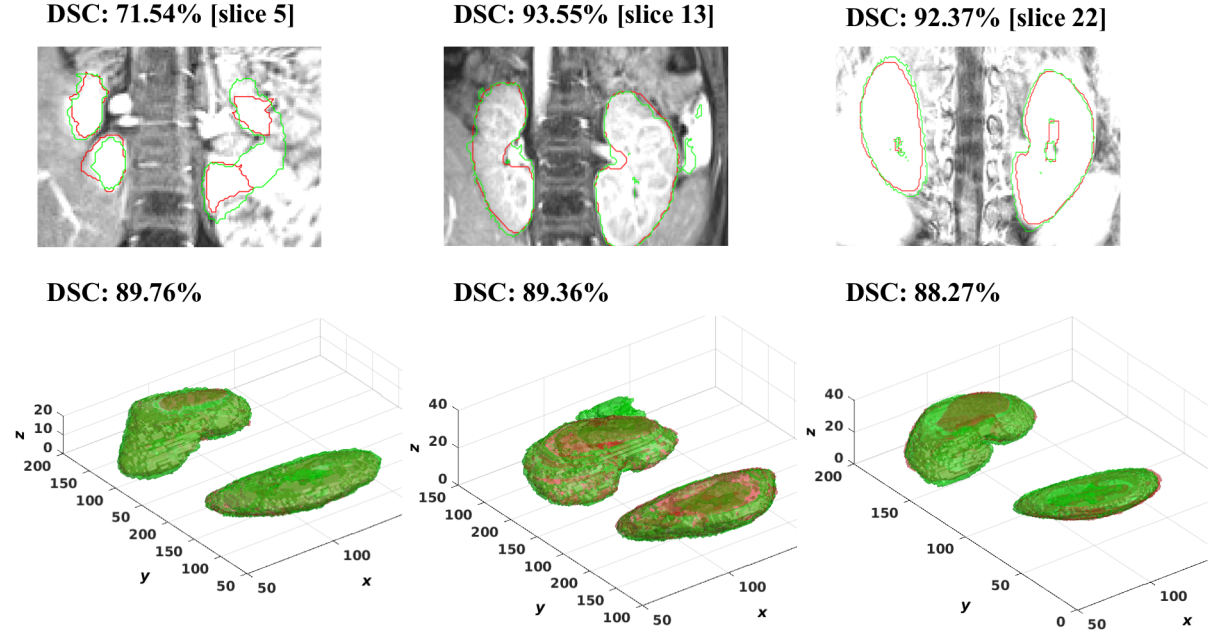


Figure 8.12: Using 3D-Method 2, a second sample batch of segmentation results in three different DCE-MRI scans (4D volumes) depicting clinically "normal" kidneys. Every column corresponds to a single DCE-MRI 4D volume. Top row displays sample slices with segmentation outcome (green) that overlap the ground-truth (red) and DSC; bottom row displays 3D reconstruction of whole kidneys segmentation that overlaps the ground-truth and DSC.

poorer in quantitative spatial accuracy. The mean DSC of $82.89 \pm 5.00\%$ demonstrates high instability, followed by a low true-positive rate as in mean precision of $73.61 \pm 7.90\%$ and a relatively stable mean recall of $95.40 \pm 1.80\%$ respectively. Although using 3D DenseNets delivers better quantitative accuracy than 3D U-Net, the approach is nonetheless inferior to 3D-Method 2, achieving mean DSC of $85.09 \pm 6.00\%$, and a mean precision and a mean recall of $82.26 \pm 6.20\%$ and $88.97 \pm 10.20\%$, respectively.

The segmentation results obtained using the "abnormal" dataset continue to demonstrate the statistical stability of 3D-Method 2, achieving a mean DSC of $86.44 \pm 3.84\%$ and a mean JI of $76.29 \pm 5.93\%$. Similarly, the mean precision and mean recall of $85.04 \pm 5.62\%$ and $88.78 \pm 8.77\%$ indicates the sustained high positive and low false-negative prediction rate. A mean HSD, albeit higher than the value obtained using the "normal" dataset, nonetheless maintains statistical stability as in $18.04 \pm 7.70mm$. The comparative baseline 3D U-Net suffers from high instability with a mean DSC of $78.94 \pm 9.10\%$, followed by an extremely poor mean precision rate that suffers in predicting true-positive tissue as in $69.55 \pm 13.40\%$. The false-negative rate is questionable with a mean quantitative accuracy of $85.08 \pm 11.30\%$. The performance of 3D DenseNets delivers better spatial overlap accuracy scores but continues to trail behind 3D-Method 2, with a mean DSC of $84.51 \pm 3.30\%$, followed by a mean precision of $82.61 \pm 5.40\%$ and a mean recall of $87.48 \pm 9.00\%$.

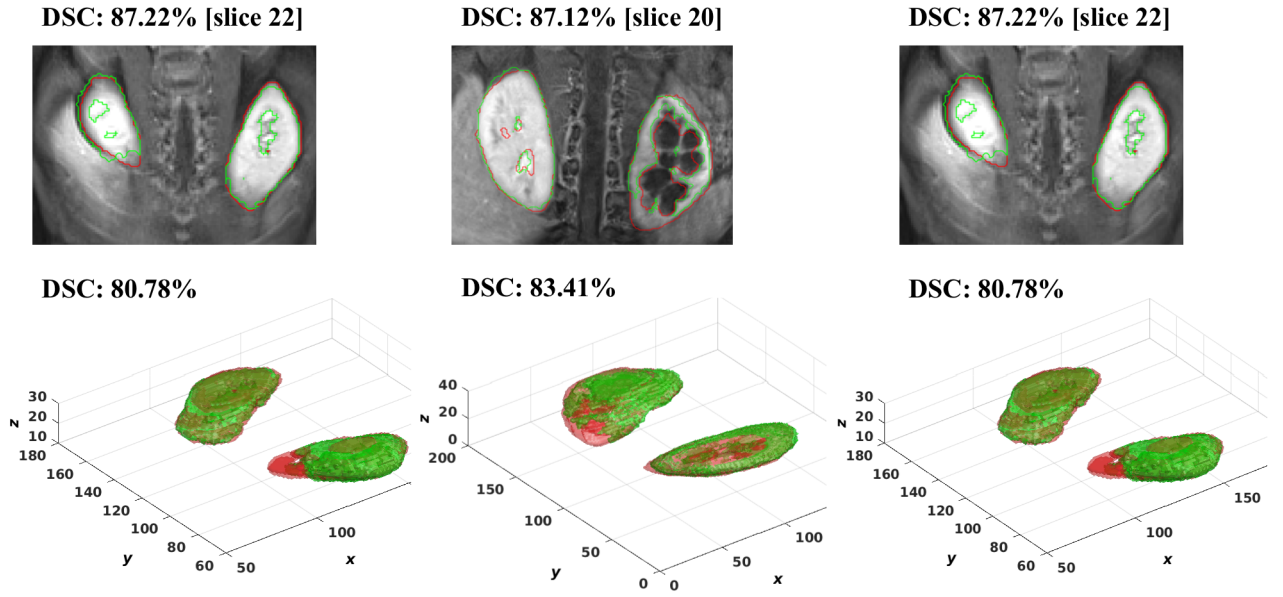


Figure 8.13: Using 3D-Method 2, a first sample batch of segmentation results in three different DCE-MRI scans (4D volumes) depicting clinically “abnormal” kidneys. Every column corresponds to a single DCE-MRI 4D volume. Top row displays sample slices with segmentation outcome (green) that overlap the ground-truth (red) and DSC; bottom row displays 3D reconstruction of whole kidneys segmentation that overlaps the ground-truth and DSC.

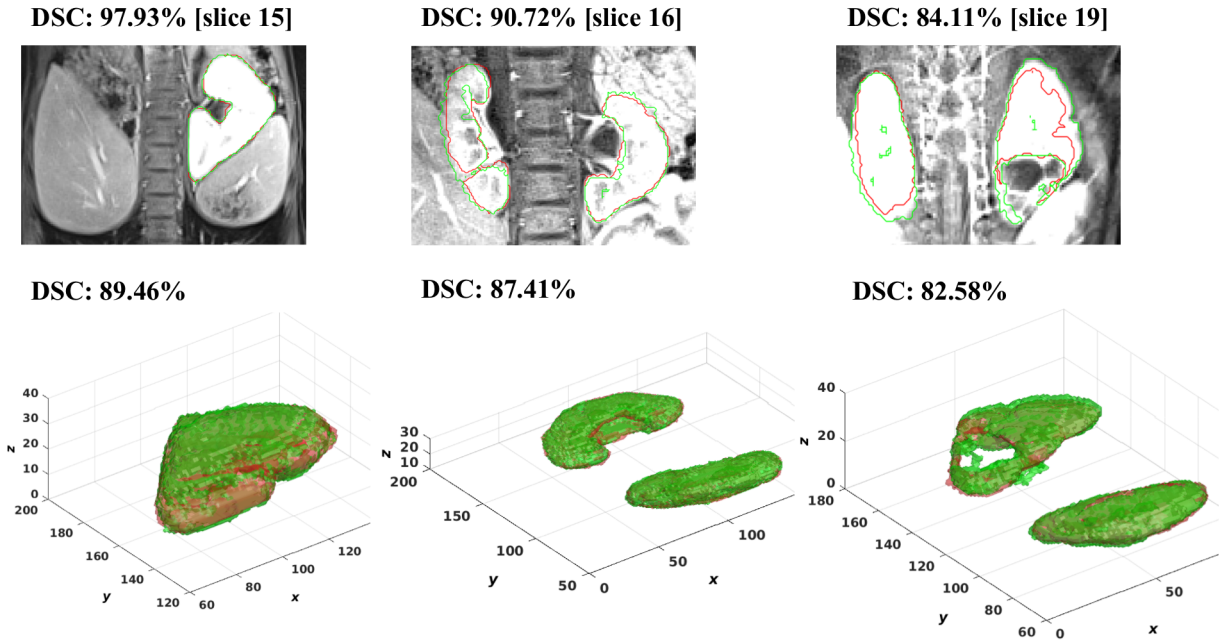


Figure 8.14: Using 3D-Method 2, a second sample batch of segmentation results in three different DCE-MRI scans (4D volumes) depicting clinically “abnormal” kidneys. Every column corresponds to a single DCE-MRI 4D volume. Top row displays sample slices with segmentation outcome (green) that overlap the ground-truth (red) and DSC; bottom row displays 3D reconstruction of whole kidneys segmentation that overlaps the ground-truth and DSC.

8.7 Future Work

With a rising need for medical image scanning, there has also been an increasing need to employ different imaging sequences to highlight particular regions or areas of interest in a scan. A segmentation model that combines different imaging sequences and modalities can overcome the limitations of singular-modality scans. For example, a combination of T1-weighted and T2-weighted sequences are obtained to detect pancreatic lesions accurately, because such pulse sequences provide necessary complementary information. In MRI breast cancer screening, contrast-enhanced imaging possesses high sensitivity in detecting breast lesions, whereas T2-weighted is effective in suppressing false lesion results [181, 182]. Also, in the case of spondylodiscitis¹², a combination of STIR (Short Tau Inversion Recovery) and T1-weighted imaging is utilised. STIR is sensitive to water, and the timing of the pulse sequence serves to minimise the signal arising in fatty tissues, and therefore only water is brightened. A comparison between conventional T1-weighted and STIR imaging can be used to evaluate the water or fat concentration within an anatomical region of interest in a scan. The abnormal signal is evident in the vertebral bodies and intervertebral disc, and so abnormal low signal on the T1-weighted sequence and abnormal high signal on the STIR image is an indication of abnormal fluid [183]. It would be of interest to develop a framework that utilises multiple modalities or sequences of the same scan. The practicality of obtaining a relevant training dataset that is sufficient for developing a deep learning model might be consolidated by building upon pre-trained singular modality models. This idea is further detailed in Chapter 9, Section 9.5.

8.8 Conclusion

In this chapter, a generalised segmentation technique is proposed with an application to other abdominal organs and muscular tissue, such as the liver, kidneys and iliopsoas muscles. A generalisation of the segmentation technique proposed in Method 2 (and 3D-Method 2) can serve as a tool for classification of anatomical measures, and therefore provide an indication or prognostic biomarker about a medical condition. Such classifications would, nonetheless, need to be founded on supporting biomedical statistical analysis. This chapter evaluates the segmentation accuracy results for liver, kidneys and iliopsoas muscles in MRI volumes, and reports close to or higher quantitative accuracies and statistical stability in comparison to state-of-the-art approaches. Furthermore, two primary motivations drive the future work building upon the PhD project detailed in this thesis: 2D and 3D deep learning method optimisation in terms of architecture and

¹²Spondylodiscitis is a combination of inflammation of intervertebral disc spaces and the inflammation of vertebrae.

computational costs; and second, extending the current or improved methodology to multiple modalities, e.g. as in MRI and US or MRI and CT. The subject regarding the *scope* of project development targeting multiple organs, muscular tissue or tumours, versus the *depth* in improving a technique targeting a single organ of high challenging variation and inter-observer error, is discussed in the next chapter.

Chapter 9

Conclusions

9.1 Introduction

The development of this PhD project began with a challenging task: segment the pancreas from one of the most readily used abdominal imaging modalities and one that is prone to high artefacts, being MRI. The pancreas is an organ of diverse structure, size and location that varies drastically from patient to patient, and associated medical conditions of diabetes and cancer is a growing interest (and concern) in the medical research and clinical community. With that said, the progression of this project has attempted to balance *scope*, *application* and *depth*. Where *scope* is concerned, the proposed approaches demonstrate generalisability in terms of application to other abdominal organs and muscular tissue, including MRI of diverse sequences and dimension (e.g. 4D data with a temporal dimension). Regarding *application*, the progression of this thesis has detailed the necessity to compute useful morphological features and biomarkers for both clinical and research purposes, in which an accurate prior organ segmentation is vital. The focus on *depth* has driven method optimisation, in which Method 2 and 3D-Method 2 (detailed in Chapter 5) are applied to two diverse MRI datasets depicting the pancreas alone.

This chapter aims to provide a summary of key challenges, motivations and achievements accomplished using the proposed methods in this PhD research project. Section 9.2 delivers a reminder of critical, technical limitations and the driving motivation for developing a system for organ segmentation in multiple modalities and multiple MRI sequences, particularly but not exclusively limited to the pancreas organ. This section examines the various drawbacks of the reported state-of-the-art and presents outperforming quantitative results using the methods that are proposed in this thesis. Section 9.3 reinstates the technical, original contributions to computer science (computer vision), and discusses the impact of these contributions to knowledge in medical image analysis. Section 9.4 compares the findings between the proposed methods in this thesis, with

a focus on segmentation performance and computational efficiency. Section 9.5 reveals the significance of employing the original contributions to knowledge into future work. Section 9.6 summarises this chapter with a conclusion reflecting upon the application of these original contributions to knowledge, and the impact in biomedical research.

9.2 Challenges and Motivation

Arguably, one of the ongoing challenges that has driven the PhD project development towards *depth* as opposed to *scope* is the high variation in pancreas segmentation accuracy in terms of spatial overlap and boundary distance error. The progression of this project has been guided by the need to develop a method that could be successfully applied to multiple MRI sequences that depict the pancreas as the object of interest. To address this first challenge, Method 2 introduces a deep learning-based loss function that effectively combines the modified Hausdorff distance metric and a sinusoidal component inspired by the cross-entropy loss function, to optimise boundary detection and improve the true-positive prediction rate. The second challenge addressed in this PhD project is the *scope* or generalisability to other objects of interest, i.e. multiple organs and muscles. Thus, this thesis has addressed the scope of segmenting the liver, iliopsoas muscles and kidneys in MRI scans acquired using entirely different sequences. The digital contrast enhancement, coupled with segmentation in 3D using continuous maximum-flow and minimum-cuts approach, continues to produce detailed boundary contouring irrespective of the target object.

As a reminder, this thesis has employed one publicly available CT dataset and two distinct MRI datasets for pancreas segmentation, from which the latter two datasets, MRI-A and MRI-B, will be taken into consideration for the purpose of this section. Moreover, a separate MRI dataset is employed for segmenting the liver and iliopsoas muscles. Furthermore, a 4D DCE-MRI dataset is split to contain clinically “normal” and “abnormal” representations of the kidneys organ.

9.2.1 Comparison to Related Work in Literature

It is noted that direct comparison with other methods in literature that utilise MRI data is difficult due to differences in scanner imaging protocols and spatial resolution. With that said, Method 2 outperforms the quantitative accuracy performance state-of-the-art research literature on MRI data. Whereas the mean of MRI-based data standard deviations in DSC from other reported publications [13, 24, 30, 130] is 7.7%, Method 2 achieves approximately two times lower a value in standard deviation of 4.6% (MRI-A) and 2.3% (MRI-B).

Moreover, employing MRI-A and MRI-B in diverse state-of-the-arts [14, 133, 171, 173, 174] demonstrated the robustness of proposed Method 2. For example, one of the drawbacks of U-Net [171], which achieved a mean Dice similarity coefficient (DSC) of 69.1% and 72.8% for MRI-A and MR-B respectively, is that network optimisation may be delayed at the stage containing the middle layers, posing a risk of ‘ignoring’ layers with abstract features that represent meaningful pancreas information. Furthermore, there is also a loss of spatial resolution in the feature maps using U-Net, posing a threat for accurate segmentation where pancreas boundary delineation is essential. In contrast, Method 2 ensures that boundary information is stored in the encoder feature maps before any sub-sampling, retaining only the max-pooling indices. When the decoder network upsamples corresponding input feature maps using the memorised max-pooling, sparse feature maps are produced and then convolved with a decoder filter to generate dense, deep feature maps containing enriched contextual information about the pancreas’ boundary. Consequently, Method 2 boosts the mean DSC performance to 84.1% and 85.7% for MRI-A and MRI-B respectively. Furthermore, the method in [14], which achieved a mean DSC 44.5% and 50.1% for MRI-A and MRI-B respectively, lacks the advantage of encoder-decoder architecture. Moreover, while the FCN [174] upsamples only once with one layer in the decoder, Method 2 has multiple upsampling layers and employs skip connections and concatenates instead of adding. The FCN, which achieved a mean DSC of 70.2% and 70.9% of MRI-A and MRI-B respectively, suffers from the drawback of not capturing local information such as the pancreas shape and structural differences. Moreover, there is the subsequent risk of network overfitting, often extremely sensitive to alternations in the pancreas’ spatial context and structure. Also, employing the Cross-entropy loss function using FCN has a drawback on predicting true-positive pixels, while the proposed Hausdorff-Sine delivers a stronger discrimination between true-positive and true-negative pixels.

The success of the generalisability or *scope* is demonstrated in the comparably high quantitative accuracies with the of state-of-the-art. For example, while a number of state-of-the-art approaches for liver segmentation achieve a mean DSC of 94% [168], 94.5% [169], 89% [170], 72.9% [171] and 93.73% [172], Method 2 both raises quantitative performance and achieves close to state-of-the-arts with a mean DSC of 94.32%. Furthermore, while the approach reported in [142] delivers a mean Jaccard index (JI) of 72.3% in psoas muscles segmentation, Method 2 boots the state-of-the-art results for segmentation composed of the psoas-muscles to a mean JI of 77.96%. Last, but certainly not least, an extension of the approach of Method 2, otherwise known as 3D-Method 2, achieves a mean DSC of 90.48% using the “normal” dataset. In contrast, the 3D U-Net performs significantly worse, reaching a mean DSC of 82.89%. While the 3D DenseNets method delivers better quantitative accuracy than 3D U-Net, the approach is nevertheless poorer to 3D-Method 2, reaching a mean DSC of 85.09%. Similarly, evaluating

the “abnormal” dataset demonstrates the robustness of 3D-Method 2. Whereas 3D U-Net trails mean DSC of 78.94 % and 3D DenseNets reaches a mean DSC of 84.51%, 3D-Method 2 outperforms with a mean DSC of 86.44%.

9.3 Impact of Original Contributions

Given the challenges that arise with MR imaging and segmenting a challenging organ such as the pancreas, the proposed technical, original contributions to computer science (computer vision) in medical image analysis, deliver: (a) a learned intensity model that enhances the contrast in image volumes to highlight the pancreas from surrounding non-pancreas tissue; (b) a hybrid energy-minimisation segmentation approach exploits edge detection to yield detailed, optimal contouring of the pancreas; (c) a post-processing stage integrates principal geometric descriptors that characterises pancreas tissue and employs radiological expert-knowledge for refined tissue classification; (d) a deep learning novel loss (error) function based on the modified Hausdorff metric and a sinusoidal component improves true-positive pancreas tissue prediction.

Moreover, the impact of the technical, original contributions to knowledge are demonstrated by:

1. Detailed boundary contouring of the pancreas in an image volume, as opposed to approximate delineation as verified by an independent senior radiologist and an independent senior radiographer. Such detailed information allows the computation of accurate pancreas curvature measures, which could facilitate stratification of clinically “normal” variations against pancreatic tumours [15] and provide enhanced information about the pancreatic curvature in cases involving ductal adenocarcinoma [10].
2. Improved application and quantitative segmentation accuracy across extended modalities such as CT volumes. Indeed, it could be useful to integrate the contributions into a clinical research tool that automatically delineates the pancreas in both MRI and CT image volumes, and computes corresponding volume and curvature measurements that serve as a guide for clinical decision-making. For example, the mean pancreas volume is reportedly 33% less in subjects with type 2 diabetes than in subjects with clinically “normal” glucose tolerance [44], while the pancreas volume is returned to near clinically “normal” after type 2 diabetes is reversed [45].
3. Robustness to diverse artefact-ridden image volumes, particularly in MR imaging. The relatively low standard deviation in DSC achieved using Method 2 with MRI-A (4.6%) and MRI-B (2.3%) compared to the lower and poor statistical stability of

[14, 133, 171, 173, 174] with MRI-A (25.2%, 17.1%, 10.2%, 10.1%, 8.5%) and MRI-B (22.7%, 18.6%, 7.5%, 11.5%, 7.7%) demonstrates the impact of the contribution to medical image segmentation in computer vision.

4. Computational efficiency in comparison to reported state-of-the-art. Unlike previous non-deep learning-based publications that require 2 to 4 hours to evaluate a single image volume [21, 25, 70] the approach described in Chapter 4 (Method 1) evaluates an image volume within 10 to 35 minutes. Moreover, the deep learning network proposed in Chapter 5 (Method 2) takes approximately 11 hours of deep network training as opposed to the approximate times of state-of-the-arts: 2 - 3 days [14, 173], 1 day [171, 174] and 21 hours [133].
5. Generalisability of the contributions extends to diverse organs and MRI sequences including the liver, iliopsoas muscles and clinically “normal” and “abnormal” kidneys. The impact of generalisability can allow the computation of multiple organ biomarkers in one single image volume, and dramatically reduce the time-consuming, laborious need for manually segmenting 3D radiological scans. It will, therefore, permit medical professional and researchers to stratify subject or patient data where correlations exist between significant changes in liver volume and curvature, and alcoholic liver disease [113]; or between kidney volume and polycystic kidney disease [117]; or between the volume of psoas muscle and predictive outcomes of treatment for bladder cancer chemotherapy [122] and ovarian cancer [123].

9.4 Comparison Between Different Proposed Methods

The contributions to knowledge as highlighted in Method 1 and 2 aim to tackle the challenges of computational efficiency and resource limitations, and the overall segmentation efficiency and robustness towards artefacts. Method 1 exploits expert radiological knowledge at the post-segmentation stage to eliminate false-positive segments identified as “pancreas”, and in doing so, provide a multi-level geometrical extraction system that is scalable and generalisable to other organs or objects of interest. Unlike subsequent methods that require an initial training stage and invest large computational resources, Method 1 requires only the training of a computationally inexpensive random forest and an artificial neural network that predicts the parameters for digital contrast enhancement. The training time for each network is comparatively tens of minutes as opposed to multiples of hours. The remainder of Method 1 can be significantly scaled at the proposed, novel refinement stage in selecting parameters for geometrical contour

(or segment) classification. Method 2, and by far 3D-Method 2, demand substantial computational resource and despite requiring a one-time deep learning training stage consisting of several hours, the testing stage (albeit minutes as opposed to hours) still need a computational head to handle the memory of the trained network. The network is scalable where training data is concerned and generalisable to other organs or target anatomical structures of interest. Furthermore, 3D-Method 2 delivers individual DSC and JI quantitative accuracies that surpass both Method 2 and Method 1. However, Method 2 achieves a higher true-positive and false-negative rate, coupled with smaller spatial overlap standard deviation. This proposed approach also delivers improved boundary delineation with a lower mean Hausdorff distance.

9.5 Employing Contributions into Future Work

The technical contributions to knowledge can integrate into a medical framework with greater scope and depth. For example, the learned intensity model for digital contrast enhancement in radiological volumes can serve as an independent module for a diverse range of MRI sequences, and other modalities that possess blurred (intensity) boundaries between the object of interest and surroundings. This independent module, combined with the hybrid energy-minimisation and structured forests approach, can amplify regions of interest in terms of intensity contrast and boundary detection, and employed into a pre-processing stage or serve as a coarse computationally inexpensive segmentation block. The modified Hausdorff based loss function can integrate into different deep learning architectures and incorporate into the formulation of new deep learning loss functions. In contrast to deep learning, one of the advantages of generating geometric descriptors per organ or object of interest (e.g. a highly inter-variable lesion) lies in the ability of the programmer to recognise the *precise* knowledge that integrates into the methodology. It can be argued that unlike hand-crafted features that are generated from machine learning algorithms (e.g. random forests) geometrical descriptors provide a concise breakdown of various shapes and structures associated with the object of interest. This information can serve as an independent source of knowledge that integrates into different algorithms, including machine or deep learning methods.

9.5.1 Extension to Multiple Modalities

As discussed previously in Chapter 8, the proposed approaches could extend to different singular modalities, such as CT or ultrasound (US), and can also integrate into methods that “fuse” extracted information from multiple modalities or imaging protocols of the same scanned region of interest. Recently, multi-sequence fusion networks as in [204] employ multiple encoders to capture features that are uniquely specific to every

different MRI sequence, and utilise a decoder to “fuse” these features and obtain high-level features for segmentation. The “fusion” process initially calibrates the low-level features again as captured from the sequence-specific encoders to emphasise both contextual information and the regions of interest, and then “fuses” the weighted features to maintain stability between “fused” and high-level features obtained from the decoder. Following a similar line of work, [205] employs a supervised image fusion method to selectively “fuse” the contextual information from multiple MRI sequences and minimise the corresponding noise. An attention block is utilised to extract useful contextual information, in which one sequence is selected as the master sequence and supervises the information selection of the remaining assistant sequences. The robustness to noise variation across differing MRI datasets (as demonstrated in Chapter 6) is an advantage that can be applied to the pre-processing stage prior to the extraction and “fusion” of image features from multiple modalities.

9.6 Conclusion

The significant challenges that limit and drive organ segmentation in 3D radiological scans, particularly MRI, are reflected in the urgency to develop effective computational tools that exploit available medical image data to its maximum capacity. This chapter has demonstrated the need to tailor the object of segmentation with an architecture that will maximise the performance in terms of spatial overlap accuracy and boundary distance error. The desire for a computational tool that can be universally applied to different major organs, muscles and tumours, from a diverse cohort of MRI sequences and other scanner imaging modalities, is of growing interest to both the clinical and biomedical research community. This thesis delivers original contributions to knowledge that are modular in the formulation and can inspire the work of upcoming computer-aided systems in medical image analysis. Furthermore, the impact of performance-based contributions to knowledge demonstrates the significance of developing systems that compute clinically useful morphological measures that provide an indication or prognostic biomarker about the severity of a medical condition, and consequently, offer enhanced treatment guidance planning for improved health and well-being.

Appendix A: Exemplars of CT and MRI Slices from Distinct Image Volumes

This appendix compares randomly selected axial slices from ten different and distinct CT abdominal volumes against randomly selected numbered slices from ten different and distinct MRI abdominal volumes, to highlight the differences in image spatial resolution and quality. Every original slice is accompanied by a duplicate slice with ground-truth (GT) contouring of the pancreas. The CT volumes are randomly selected from the publicly available dataset CT-NIH (<http://dx.doi.org/10.7937/K9/TCIA.2016.tNB1kqBU>) and the MRI volumes are randomly selected from the datasets MRI-A and MRI-B described in Chapter 6.

**Original CT Slice
in Distinct Volume**

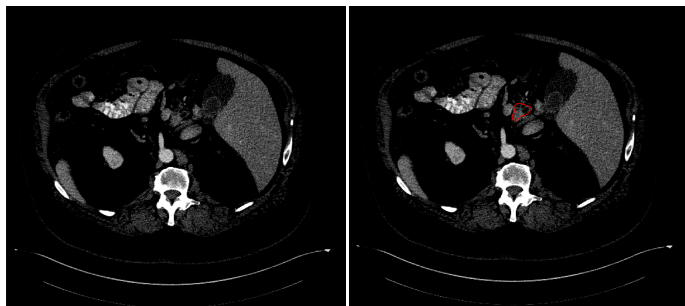
**Pancreas Contour
(GT) in CT Slice**

**Original MRI Slice
in Distinct Volume**

**Pancreas Contour
(GT) in MRI Slice**

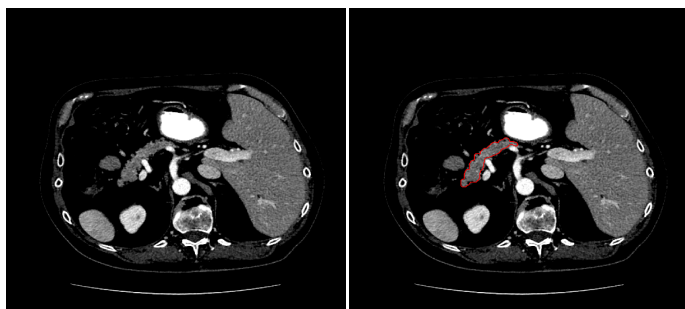
116

116 - GT



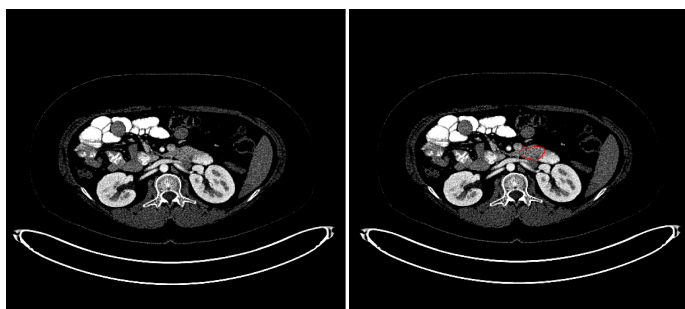
95

95 - GT



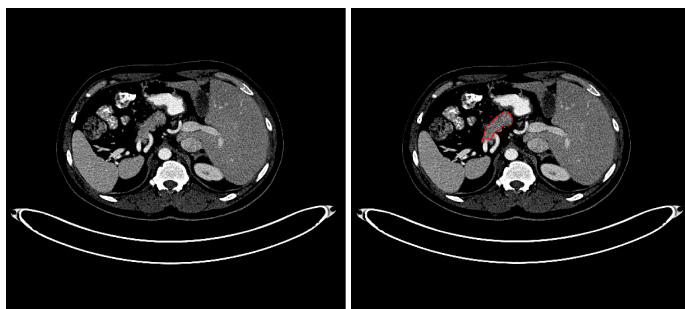
149

149 - GT



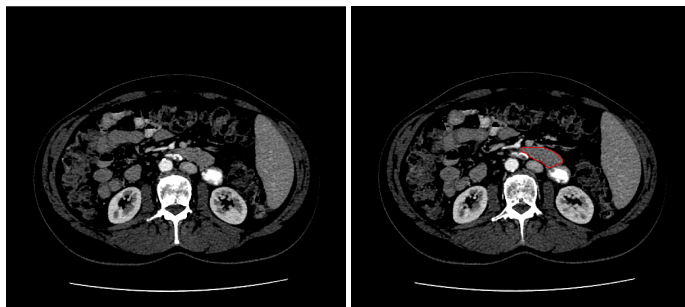
86

86 - GT



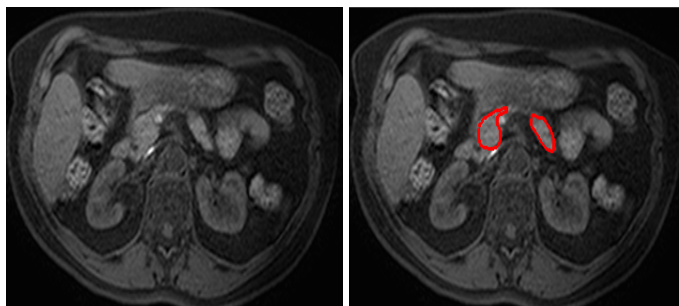
158

158 - GT



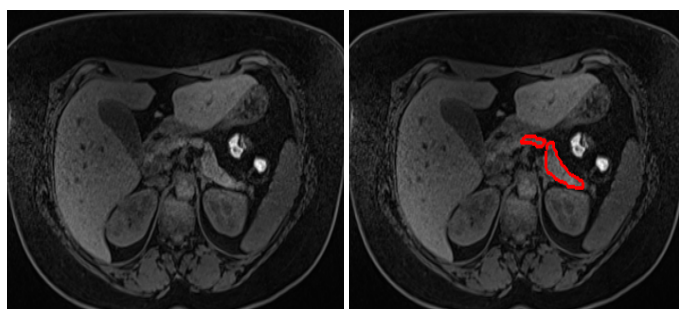
13

13 - GT



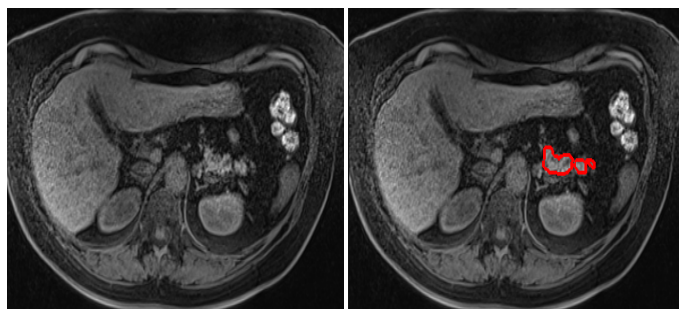
49

49 - GT



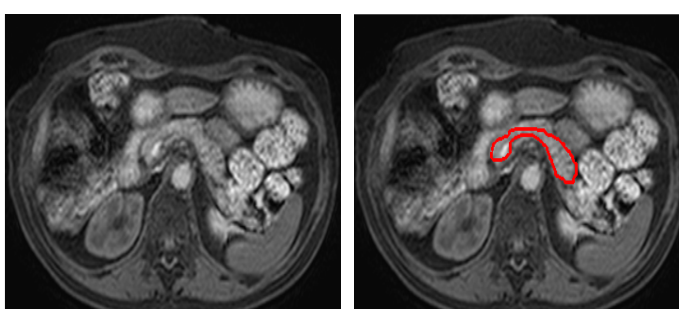
44

44 - GT



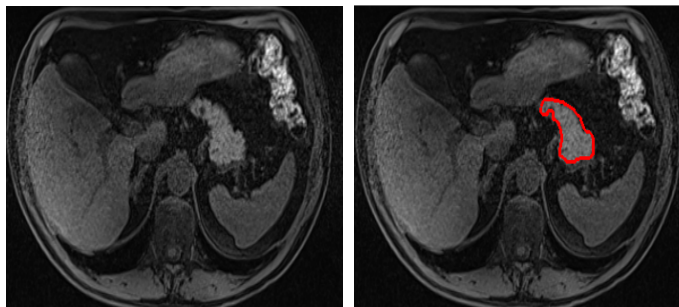
17

17 - GT



59

59 - GT



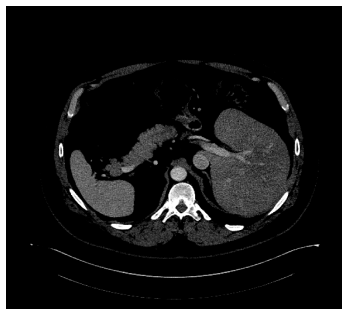
**Original CT Slice
in Distinct Volume**

**Pancreas Contour
(GT) in CT Slice**

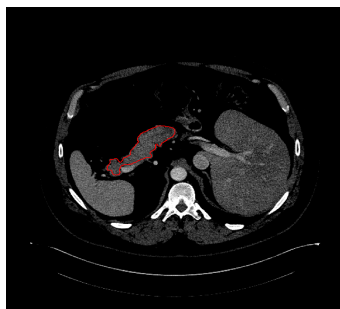
**Original MRI Slice
in Distinct Volume**

**Pancreas Contour
(GT) in MRI Slice**

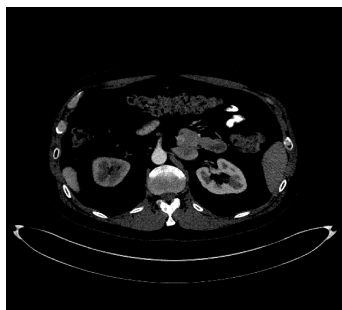
100



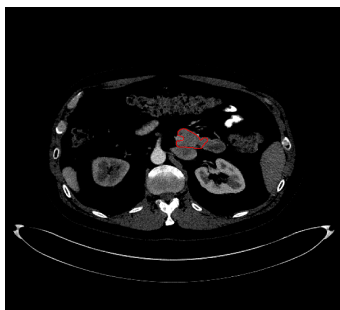
100 - GT



175



175 - GT



60



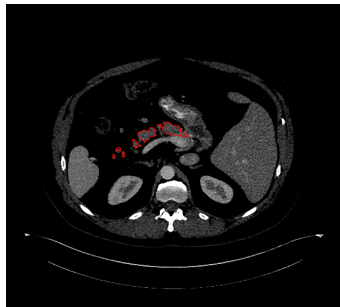
60 - GT



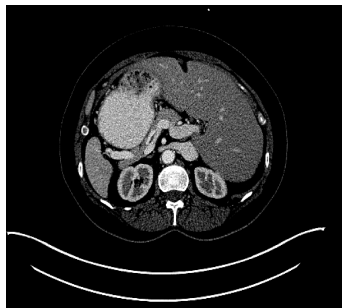
133



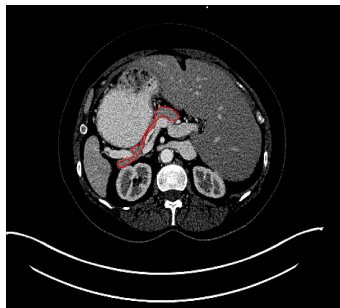
133 - GT



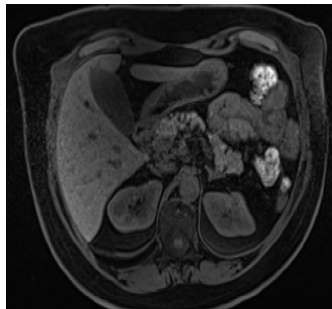
94



94 - GT



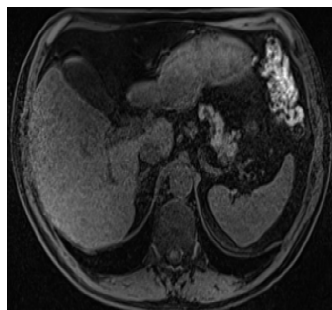
37



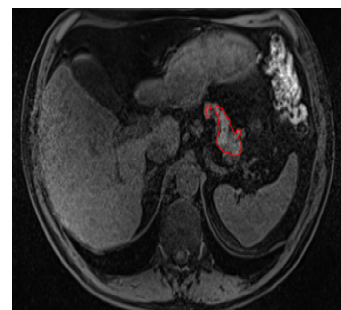
37 - GT



62



62 - GT



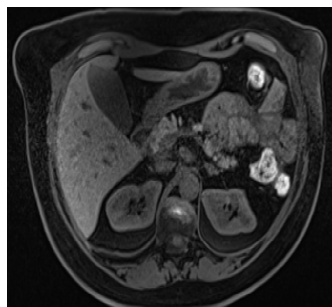
30



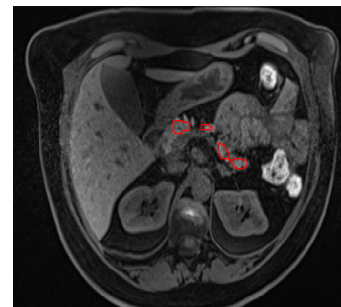
30 - GT



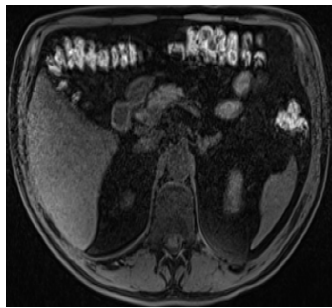
34



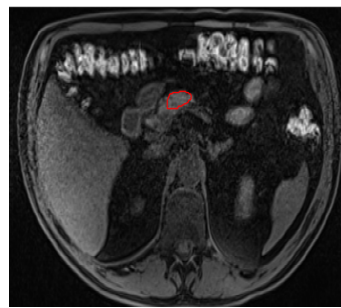
34 - GT



44



44 - GT



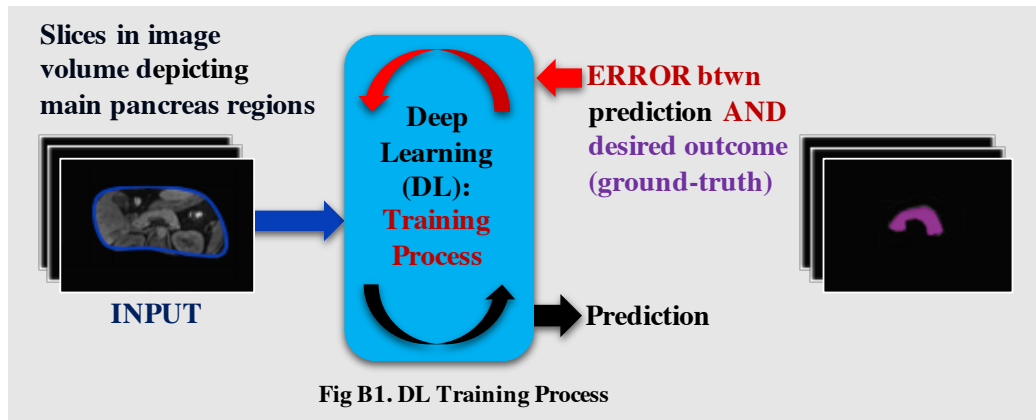
Appendix B: Deep Learning Network Development and Convolutional Neural Network Architecture

This appendix provides a “bird’s eye view” description about the training and testing process of a deep learning network as shown in Fig. B1 and Fig. B2 on page 154, and the main objective of each layer in a general form of a convolutional neural network as shown in Fig. B3 on page 155.

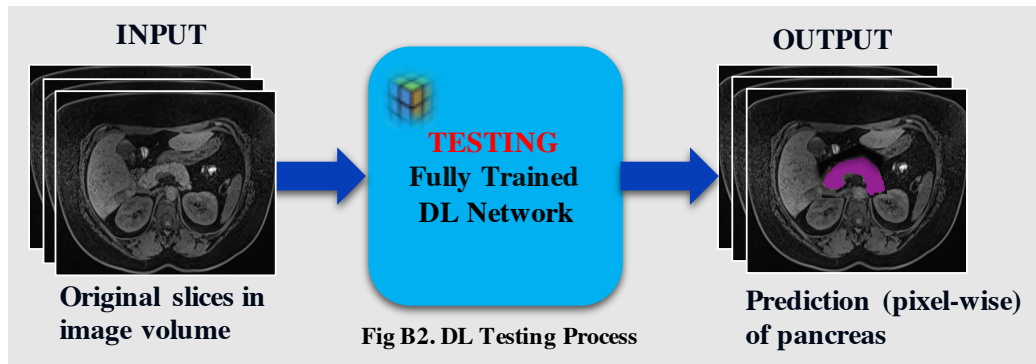
A **deep learning (DL)** network is trained using multiple image volumes and respective annotated ground-truth of the pancreas.

The aim of the **training process**, as shown in Fig. B1, is to optimise the DL network to ‘recognise’ distinguished features and boundaries of the pancreas.

The DL network is optimised by iteratively minimising the error between the output prediction and the actual desired outcome (ground-truth) using a loss function through a technique called **backpropagation**.



At the **testing stage**, an image volume is processed through the fully trained DL network, as shown in Fig. B2, which in turn performs pixel-wise or voxel-wise prediction to output a binarised sequence of 2D slices or a 3D volume differentiated by “pancreas” and “non-pancreas” labels.



Backpropagation is the process where the DL network is optimised by updating the convolutional layers’ filters and minimising the overall loss between the network prediction and desired outcome (ground-truth). This process employs an initial learning rate parameter that changes via a “learning rate drop period” and a “learning rate drop factor”. The learning rate drop period is the number of epochs for dropping the learning rate and the learning rate drop factor is a multiplicative factor that is applied to the learning rate every time a certain number of epochs passes. An epoch refers to a full pass of the entire training data, whereas an iteration number refers to a full pass of the mini-batch (a subset of the entire training data).

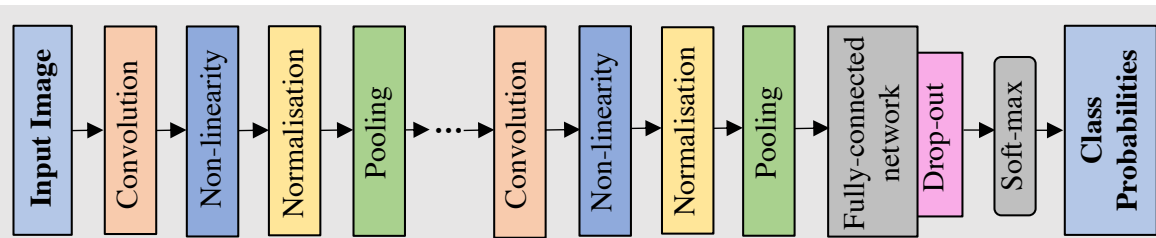


Fig B3. Overview of layers in Convolutional Neural Network (CNN) architecture

A **convolutional layer** produces feature maps that possess information or ‘features’ extracted from an input image or previous feature map.

- A $n \times n$ *local receptive field* convolves across a given input, of $p \times q$ size (where $p, q > n$) using a stride, s (where $s \geq 1$). This receptive field occupies a set of $n \times n$ matrices or shared *weights*, also known as a *filter*. The outcome of this convolution is known as a *feature map* or *activation map*.

A **non-linearity layer** increases non-linearity during DL training.

The ReLU (Rectified Linear Units) is a popular function of choice for training.

- ReLU layer has no filter and does not change the size of the input feature map. Taking on board a feature map that possesses positive, negative and zero values, the ReLU layer will set all negative values to 0.

A **normalisation layer** can accelerate optimisation for faster convergence.

- A normalisation layer performs scaling on data in the feature maps to suitable intervals, consequently removing potential bias. Examples of the normalisation layer include batch normalisation, brightness normalisation and subtraction normalisation.

A **pooling layer** serves to simplify the information in the output from the previous layer. A pooling operation requires no weights.

- The most common form is a max-pooling layer with a small $n \times n$ region, e.g. 2×2 and a stride of 2. The max-pooling operation down-samples by taking the maximum value from four elements in every 2×2 region.

A **fully-connected layer** represents a Multilayer Perceptron (MLP) network.

- A MLP network is a type of artificial neural network (ANN), which is one-dimensional and consists of three or more layers: an input and an output layer, and one or more hidden layers. All the inputs from one layer are connected to every feature map value of the next (hidden) layer.

A **drop-out** method prevents overfitting and improves network convergence: reduces the complex interactions between input values, enabling the network optimisation of features that are meaningful towards recognising the pancreas.

A **soft-max layer** is a K-way soft-max function which produces a probability distribution over K-many classes. This function’s output are class probabilities of the input image across 2 classes: “pancreas” and “non-pancreas”.

Appendix C: Qualitative Expert Review of Pancreas Contouring Outcomes

This appendix provides the qualitative feedback of an independent senior radiologist who analysed the pancreas contouring outcome in five different and distinct MRI volumes from the datasets MRI-A and MRI-B described in Chapter 6 using manual segmentation versus the proposed segmentation approach of Method 2 (Chapter 5).

Review: Contouring Using Manual Technique vs. Proposed Approach

Please find below the full review by an independent senior radiologist who compared the pancreas contouring in five different MRI volumes obtained using a manual technique (referred as red) with the proposed approach of Method 2 (referred as green).

The name of each image volume has been replaced with [Case number (MRI dataset type)] to correspond to the datasets described in Chapter 6.

"I carefully looked at all the slices you sent to me, and after each group of them, I give you my opinion. I hope that it will be of benefit to you.

In [Case 1 (MRI-B)] Slice 55 and 56 with green denotes the part of the lining vein, that is, the green line behind a pancreas exits beyond its outline. On this slice, the pancreas countertops with the red line are better marked, although in the front of the ophthalmic area is a shortened part of the pancreas parenchyma.

In [Case 2 (MRI-B)] with green beans, the pancreas contours are best marked, such as protrusions, as well as dents or incisions. Most of the surface of the pancreas is covered. While the marked with a red line does not fully follow the contours of the pancreas, it cuts them and rounds the incisions or dents, rounds it around.

In [Case 3 (MRI-B)] on Slice 39 with a green line is a shortened part of the pancreas contour, a projection marked with a red line. But it is therefore missed to signify a projection in the back of the pancreas in the area of the body, marked with a red line. On this slice, the overall accuracy is almost equal, with a slight advantage for the green line. Slice 45 has a big mistake in marking the green line, a part that does not belong to the pancreas is indicated. The red dot mark is more accurate.

In [Case 1 (MRI-A)], the head of the pancreas in Slice 1, 2, 3, 4, 5 and 6 is not marked, either with the red or green line. On Slices 7 and 8, a small portion of the pancreas head is indicated only with a red line, and on a Slice 9, 10 and 11, a small part of the pancreas head is indicated by both lines.

In [Case 2 (MRI-A)] on the Slice 28, 29 and 30 with a green line are marked parts of the back wall of the stomach. In this series, it is generally better to mark the red line.

Characteristic for all lines and for all series is that the green line, however, better limits the pancreas to its protrusions and recesses, while the red line tends to level the contours of the pancreas, that is, it does not fully follow the protrusions and the dentures or the incisions of the pancreas contour".

Furthermore, pages 158 – 167 present a table for each of the five analysed MRI volumes. For every slice in each MRI volume, the following two areas have been reviewed in terms of (1) overall accuracy and (2) the protrusions and indentations of the pancreas' contouring. The senior radiologist has placed an 'X' in the relevant box of each slice that they interpret provides the best overall accuracy and contouring.

Image Case (Folder) Name: Case 1 (MRI-B)

Contour visibility: Slices 40 - 66

Slice Number	Overall accuracy		Protrusions and indentations of the contour(s)	
Slices 40 - 66	Green (a)	Red (b)	Green (a)	Red (b)
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				
21				
22				
23				
24				
25				
26				
27				
28				
29				
30				
31				
32				
33				
34				
35				
36				
37				
38				
39				
40	X		X	
41	X		X	
42	X		X	
43	X		X	
44	X		X	
45	X		X	
46	X		X	
47	X		X	
48	X		X	

49	X		X	
50	X			
51	X			
52	X			
53	X			
54	X			
55		X		X
56		X		X
57		X		X
58	X		X	
59	X		X	
60	X		X	
61	X		X	
62	X		X	
63	X		X	
64	X		X	
65	X		X	
66		X		X
67				
68				
69				
70				
71				
72				
73				
74				
75				
76				
77				
78				
79				
80				

Image Case (Folder) Name: Case 2 (MRI-B)

Contour visibility: Slices 30 - 50

Slice Number	Overall accuracy		Protrusions and indentations of the contour(s)	
Slices 30 - 50	Green (a)	Red (b)	Green (a)	Red (b)
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				
21				
22				
23				
24				
25				
26				
27				
28				
29				
30	X		X	
31	X		X	
32	X		X	
33	X		X	
34	X		X	
35	X		X	
36	X		X	
37	X		X	
38	X		X	
39	X		X	
40	X		X	
41	X		X	
42	X		X	
43	X		X	
44	X		X	
45	X		X	
46	X		X	
47	X		X	
48	X		X	

49	X			
50	X		X	
51				
52				
53				
54				
55				
56				
57				
58				
59				
60				
61				
62				
63				
64				
65				
66				
67				
68				
69				
70				
71				
72				
73				
74				
75				
76				
77				
78				
79				
80				

Image Case (Folder) Name: Case 3 (MRI-B)

Contour visibility: Slices 30 - 52

Slice Number	Overall accuracy		Protrusions and indentations of the contour(s)	
Slices 30 - 52	Green (a)	Red (b)	Green (a)	Red (b)
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				
21				
22				
23				
24				
25				
26				
27				
28				
29				
30	X		X	
31	X		X	
32	X		X	
33	X		X	
34	X		X	
35	X		X	
36		X		X
37	X			
38	X			
39	X			X
40	X		X	
41	X		X	
42	X		X	
43	X		X	
44	X		X	
45		X		X
46	X		X	
47	X		X	
48	X		X	

49	X		X	
50				
51				
52				
53				
54				
55				
56				
57				
58				
59				
60				
61				
62				
63				
64				
65				
66				
67				
68				
69				
70				
71				
72				
73				
74				
75				
76				
77				
78				
79				
80				

Image Case (Folder) Name: Case 1 (MRI-A)

Contour visibility: Slices 7 - 32

Slice Number	Overall accuracy		Protrusions and indentations of the contour(s)	
Slices 7 - 32	Green (a)	Red (b)	Green (a)	Red (b)
1				
2				
3				
4				
5				
6				
7		X		X
8		X		X
9	X			X
10	X		X	
11		X		X
12		X		X
13		X		X
14		X		X
15		X		X
16		X		X
17	X		X	
18	X		X	
19	X		X	
20	X		X	
21	X		X	
22	X		X	
23	X		X	
24	X		X	
25	X		X	
26	X		X	
27	X		X	
28	X		X	
29		X		X
30		X		X
31		X		X
32		X		X
33				
34				
35				
36				
37				
38				
39				
40				
41				
42				
43				
44				
45				
46				
47				
48				

49				
50				

Image Case (Folder) Name: Case 2 (MRI-A)

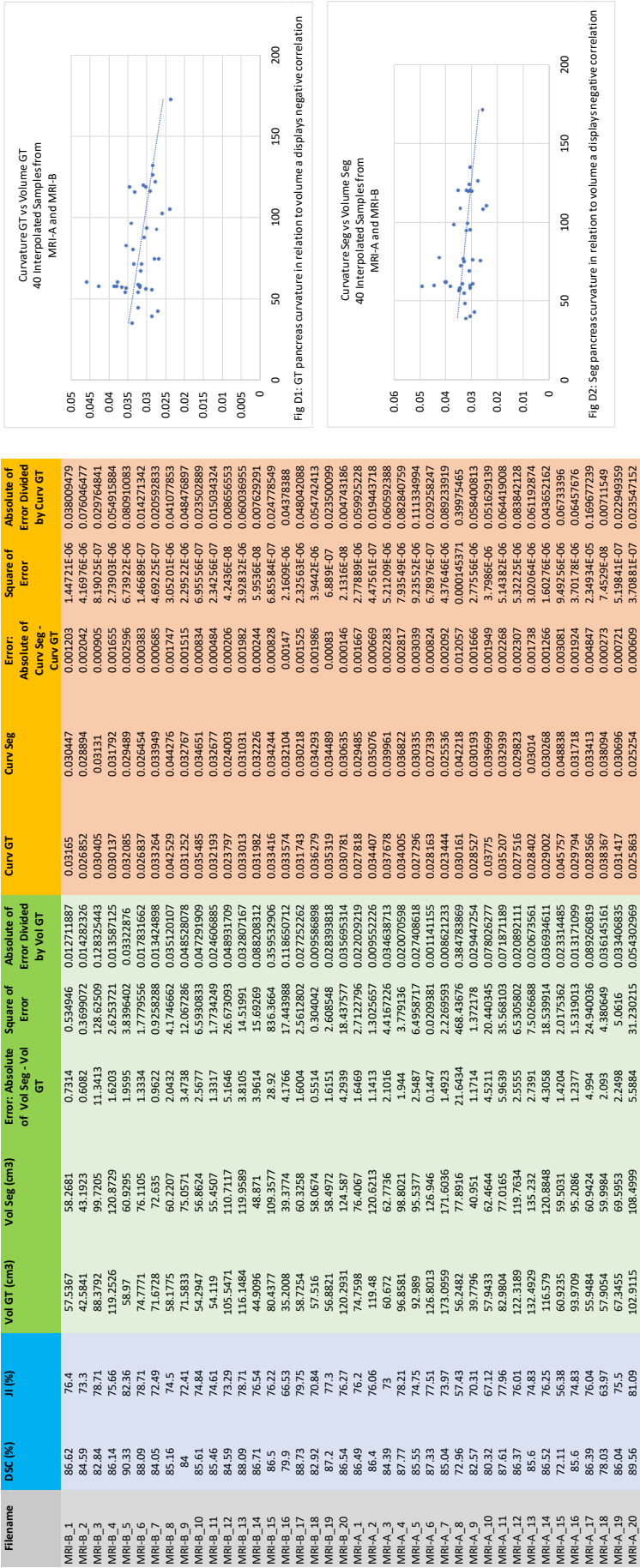
Contour visibility: Slices 12 - 32

Slice Number	Overall accuracy		Protrusions and indentations of the contour(s)	
Slices 12 - 32	Green (a)	Red (b)	Green (a)	Red (b)
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12	X		X	
13	X		X	
14	X		X	
15		X		X
16		X		X
17		X		X
18		X		X
19		X		X
20		X		X
21		X		X
22		X		X
23		X		X
24		X		X
25		X		X
26		X		X
27		X		X
28		X		X
29		X		X
30		X		X
31		X		X
32		X		X
33				
34				
35				
36				
37				
38				
39				
40				
41				
42				
43				
44				
45				
46				
47				
48				

49				
50				

Appendix D: Quantitative Morphological Features of Pancreas Volume and Curvature

This appendix lists forty individual DSC and JI scores of pancreas segmentation using the MRI volumes from datasets MRI-A and MRI-B (Chapter 6) based on Method 2 (Chapter 5). The respective ground-truth (GT) and automatic segmentation (Seg) results for quantitative morphological features, including volume and curvature, are also provided. This appendix also provides the MAE (mean absolute error), RMSE (root mean square error) and MAP (mean absolute percentage error) computations for both volume and curvature. Furthermore, two figures compare the relationship between pancreas curvature and volume for both GT (Fig. D1) and Seg (Fig. D2) results.



MEAN ACC (%)	84.918	74.1715
STD (%)	3.729261	5.355596395
ACC = Accuracy Measure (DSC or JI)		
DSC = Dice Similarity Coefficient		
JI = Jaccard Index		

Vol MAE:	3.83924	6.6076124
Vol RMSE :		
Vol MAP (%) :		5.384225449
Vol MAP - STD (%) :		7.84282159
Mean Vol GT +/- STD:	80.47527	30.975499
Mean Vol Seg +/- STD:	83.742905	30.937992
Vol GT = Ground-truth Volume		
Vol Seg = Automated Segmentation Volume		

Curv MAE:	0.001784075	0.002607762
Curv RMSE:		
Curv MAP (%) :		5.723089758
Curv MAP - STD (%) :		6.350304187
Mean Curv GT +/- STD:	0.031793325	0.004586267
Mean Curv Seg +/- STD:	0.0326949	0.004992605
Curv GT = Ground-truth Curvature		
Curv Seg = Automated Segmentation Curvature		

STD = Standard Deviation
MAE = Mean Absolute Error
RMSE = Root Mean Square Error
MAP = Mean Absolute Percentage

Appendix E: Number of Combinations for Contours with Non-infinity Gradient

With reference to Section 4.4.3 Level 3: Centre Landmarks of Distinct Contours (Chapter 4), this appendix describes the formulae used to derive the number of combinations for every two, three and four contour combination where the gradient between two contour centres is non-infinity.

Referring to Section 4.4.3 (Chapter 4), the total number of possible combinations of T “pancreas” contours from M available contours is computed, as denoted by C_{TM} and defined as:

$$C_{TM} = \frac{M!}{T!(M-T)!}$$

where,

M is the total number of separate contours in a slice;

N is the total number of vertical lines that have two contours, i.e., where the gradient between respective centre points is infinity;

$T = (M - N)$ is the number of “pancreas” contours to be retained after processing.

Analysing two and three contour combinations The number of combinations to be considered, in which the gradient between two contour centre points is non-infinity, $C_{TM}^{2,3}$, can be described as:

$$C_{TM}^{2,3} = C_{TM} - \frac{N(M-2)!}{(T-2)!(M-T)!}$$

Analysing four-contour combinations The number of combinations to be considered, in which the gradient between two contour centre points is non-infinity, C_{TM}^4 , can be described as:

$$C_{TM}^4 = C_{TM} - \left(\frac{N(M-2)!}{(T-2)!(M-T)!} + \frac{N(N-1)}{2} \right)$$

Example For example, considering Fig. E1, if $M = 5$, $N = 2$, thus $T = 3$, then $C_{3,5} = 10$ different combinations of three “pancreas” contours from five available contours in a slice (2D image) from a test image volume.

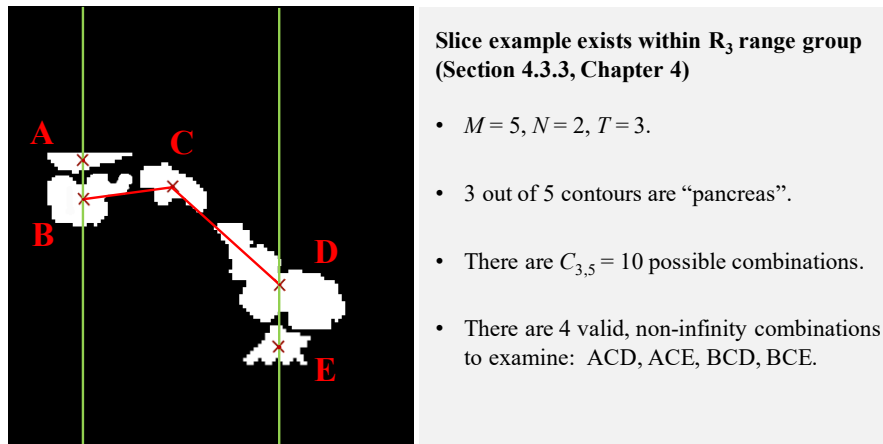


Fig E1. Visual example of a three-contour combination in a test image slice (2D).

Therefore,

$$C_{3,5} = \frac{5!}{3!(5-3)!} = 10$$

Hence,

$$C_{3,5}^3 = 10 - \frac{2(5-2)!}{(3-2)!(5-3)!} = 4$$

There are 4 valid combinations to examine during processing as shown in Fig. E1.

Appendix F: Example of Level 3 Refinement

With reference to Section 4.4.3 Level 3: Centre Landmarks of Distinct Contours (Chapter 4), this appendix provides an example of employing the “ground-truth characteristics” set and respective “segmentation characteristics” group relating to R_3 to remove segmented contours identified as “non-pancreas”.

Let us consider the “ground-truth characteristics” set represented by f_p^i where i refers to the i -th slice range group, R_i , and p refers to the number of contours. Therefore, f_3^3 refers to R_3 and a three-contour combination:

$$\begin{aligned}
f_3^3 &= \{(A, B, C, g_{AB}, g_{BC}, g_{AC}, d_{AB}, d_{AC}, d_{BC}, A/B, A/C, B/C)_1 \\
&= (A, B, C, g_{AB}, g_{BC}, g_{AC}, d_{AB}, d_{AC}, d_{BC}, A/B, A/C, B/C)_2 \\
&= (A, B, C, g_{AB}, g_{BC}, g_{AC}, d_{AB}, d_{AC}, d_{BC}, A/B, A/C, B/C)_3 \\
&\quad \dots \\
&= (A, B, C, g_{AB}, g_{BC}, g_{AC}, d_{AB}, d_{AC}, d_{BC}, A/B, A/C, B/C)_n\}
\end{aligned}$$

where n ($= 36$ in this example) is a chosen number of groups in a set that is based on k -medoids clustering of these measurements in the training dataset. A, B, C refer to the areas of individual contours, g_{AB}, g_{BC}, g_{AC} and d_{AB}, d_{AC}, d_{BC} refer to the gradient and Euclidean distance between centre points of paired contours. Last, $A/B, A/C, B/C$ refer to the ratio of areas of paired contours.

Next, the total number of possible combinations of T “pancreas” contours from M available contours is computed, as denoted by C_{TM} and defined as:

$$C_{TM} = \frac{M!}{T!(M-T)!}$$

For every $C_{TM}(j)$ -th combination of contours, where $j \in \mathbb{Z} : j \in [1, C_{TM}]$, the gradients of all possible paired contours are calculated, in addition to respective distances and ratios.

As shown in Figure F1, this example has $M = 5$ and $T = 3$, such that $C_{3,5} = 10$ different combinations of three contours from five contours (where two are presumed invalid). Therefore:

$$\begin{aligned}
C_{3,5}(1) &= A, B, C \\
C_{3,5}(2) &= A, B, D \\
C_{3,5}(3) &= A, B, E \\
C_{3,5}(4) &= A, C, D \\
C_{3,5}(5) &= A, C, E \\
C_{3,5}(6) &= A, D, E \\
C_{3,5}(7) &= B, C, D \\
C_{3,5}(8) &= B, C, E \\
C_{3,5}(9) &= B, D, E \\
C_{3,5}(10) &= C, D, E
\end{aligned}$$

A number of combinations will be irrelevant since the gradient between two centres is infinity, and thus can be discarded. Hence, for every remaining (relevant) three-contour combination, a “segmentation characteristics” or “sc” group is generated, consisting

of three gradients, areas, Euclidean distances and ratios-of-areas. Therefore, in this example:

$$\begin{aligned}
C_{3,5}(4) - \text{"sc"} &= (A, C, D, g_{AC}, g_{AD}, g_{CD}, d_{AC}, d_{AD}, d_{CD}, A/C, A/D, C/D) \\
C_{3,5}(5) - \text{"sc"} &= (A, C, E, g_{AC}, g_{AE}, g_{CE}, d_{AC}, d_{AE}, d_{CE}, A/C, A/E, C/E) \\
C_{3,5}(7) - \text{"sc"} &= (B, C, D, g_{BC}, g_{BD}, g_{CD}, d_{BC}, d_{BD}, d_{CD}, B/C, B/D, C/D) \\
C_{3,5}(8) - \text{"sc"} &= (B, C, E, g_{BC}, g_{BE}, g_{CE}, d_{BC}, d_{BE}, d_{CE}, B/C, B/E, C/E)
\end{aligned}$$

Each group is simultaneously evaluated against a set of “ground-truth characteristics” from the same slice range group and T contours of interest. Here, set f_3^3 is used.

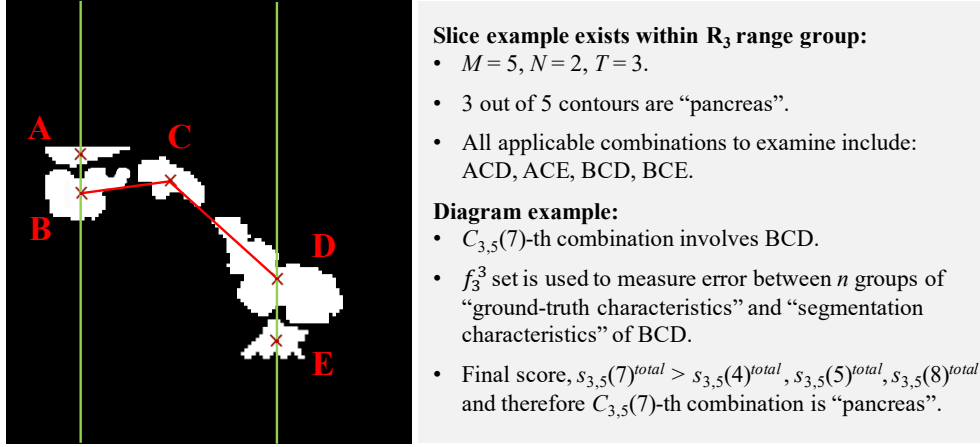


Fig F1. Visual example of a three-contour combination in a test image slice (2D).

A method of elimination follows: if the mean error between the $C_{TM}(j)$ -th “segmentation characteristics” group and a “ground-truth characteristics” group is equal or falls below threshold, t_p , then this combination is assigned a point score of 1, otherwise 0. Once all measures of error have been exhausted, the combination with the highest score is retained as “pancreas”, and the others discarded. The equation below describes the process of assigning a score to the j -th combination, $s_{TM}(j)^k$, where $k \in \mathbb{Z} : k \in [1, n]$.

$$s_{TM}(j)^k = \begin{cases} 1 & \text{if } 0 \leq E_{f_p^i}^k \leq t_p \\ 0 & \text{otherwise} \end{cases}$$

where $E_{f_p^i}^k$ is the mean error between the $C_{TM}(j)$ -th combination “segmentation characteristics” and $(f_p^i)_k$ which is the k -th “ground-truth characteristics”. The final score, $s_{TM}(j)^{total}$ can be described using the equation below:

$$s_{TM}(j)^{total} = \sum_{k=1}^n s_{TM}(j)^k$$

Therefore, in this example:

$$s_{3,5}(4)^{total} = 10 \Rightarrow A, C, D$$

$$s_{3,5}(5)^{total} = 2 \Rightarrow A, C, E$$

$$s_{3,5}(7)^{total} = 19 \Rightarrow B, C, D$$

$$s_{3,5}(8)^{total} = 2 \Rightarrow B, C, E$$

Thus, B, C, D is chosen as the best three-contour combination “pancreas”.

Appendix G: Derivation of Modified Hausdorff Loss

This appendix provides extends upon Section 4.5 Integrated Hausdorff-Sine Loss Function (Chapter 5), providing a breakdown of the derivation of gradients for the first part (that is L_{mhd}) in the proposed loss function.

A brief reminder from Section 4.5. Let T_H and Y_H represent the ground-truth (target) and network boundary (or edge) predictions, respectively; and where $T_H, Y_H \subset \mathbb{R}^n$ such that $|T_H|, |Y_H| < \infty$. Furthermore, t_j and $y_j \in \{0, 1\}$ are indexed pixel values in T_H and Y_H respectively, which can subsequently, be viewed as boundary points. From here, it is possible to define the minimum of the Euclidean (2D) distance between a point t_j and set of points, Y_H as: $s(t_j, Y_H) = \min_{y_j \in Y_H} \|t_j - y_j\|$, and similarly, the minimum distance between a point y_j and set of points, T_H , can be defined as: $s(y_j, T_H) = \min_{t_j \in T_H} \|y_j - t_j\|$.

Whereas the Hausdorff distance (HSD) is defined as:

$$HSD = \max \left\{ \max_{y_j \in Y_H} s(y_j, T_H), \max_{t_j \in T_H} s(t_j, Y_H) \right\}$$

The modified Hausdorff distance is less sensitive to outliers and thus the loss, L_{mhd} , can be described as:

$$L_{mhd} = \max \left\{ \frac{1}{|Y_H|} \sum_{y_j \in Y_H} s(y_j, T_H), \frac{1}{|T_H|} \sum_{t_j \in T_H} s(t_j, Y_H) \right\}$$

Thus, computing the gradient yields:

$$\frac{\partial L_{mhd}}{\partial Y_H} = \begin{cases} \frac{\partial}{\partial Y_H} \left(\frac{1}{|Y_H|} \sum_{y_j \in Y_H} s(y_j, T_H) \right) & \text{if } \frac{1}{|Y_H|} \sum_{y_j \in Y_H} s(y_j, T_H) > \frac{1}{|T_H|} \sum_{t_j \in T_H} s(t_j, Y_H) \\ \frac{\partial}{\partial Y_H} \left(\frac{1}{|T_H|} \sum_{t_j \in T_H} s(t_j, Y_H) \right) & \text{if } \frac{1}{|Y_H|} \sum_{y_j \in Y_H} s(y_j, T_H) < \frac{1}{|T_H|} \sum_{t_j \in T_H} s(t_j, Y_H) \\ \text{undefined} & \text{if } \frac{1}{|Y_H|} \sum_{y_j \in Y_H} s(y_j, T_H) = \frac{1}{|T_H|} \sum_{t_j \in T_H} s(t_j, Y_H) \end{cases}$$

where,

$$\frac{\partial}{\partial Y_H} \left(\frac{1}{|Y_H|} \sum_{y_j \in Y_H} s(y_j, T_H) \right) = \frac{1}{|Y_H|} \sum_{y_j \in Y_H} \min_{t_j \in T_H} \left(\frac{\sum_{k=1}^2 (y_j - t_j)_k}{\|y_j - t_j\|} \right)$$

and

$$\frac{\partial}{\partial Y_H} \left(\frac{1}{|T_H|} \sum_{t_j \in T_H} s(t_j, Y_H) \right) = \frac{1}{|T_H|} \sum_{t_j \in T_H} \min_{y_j \in Y_H} \left(\frac{\sum_{k=1}^2 (y_j - t_j)_k}{\|t_j - y_j\|} \right)$$

Breakdown of derivation

Check point 1: the chain rule states that, $\frac{dy}{dx} = \frac{dy}{du} \times \frac{du}{dx}$. Therefore, if $y = f(x)^n$ and $u = f(x)$ we can summarise that,

$$\frac{d}{dx} f(x)^n = n f(x)^{n-1} \frac{d}{dx} f(x)$$

Check point 2: the derivation of the summation operator states that,

$$\begin{aligned} \frac{d}{dx} \sum_{i=1}^n f_i(x) &= \frac{d}{dx} (f_1(x) + \dots + f_n(x)) \\ &= \frac{df_1}{dx} + \dots + \frac{df_n}{dx} \\ &= \sum_{i=1}^n \frac{df_i}{dx} \end{aligned}$$

A) Let us express the Euclidean norm between point y_j and t_j in 2D space as:

$$\|y_j - t_j\| = \sqrt{\sum_{k=1}^2 (y_{j_k} - t_{j_k})^2}$$

Let us consider the partial derivative of $\|y_j - t_j\|$ with respect to y_j . Thus,

$$\begin{aligned} \frac{\partial}{\partial y_j} \left(\sqrt{\sum_{k=1}^2 (y_{j_k} - t_{j_k})^2} \right) &= \frac{1}{2} \times \frac{1}{\sqrt{\sum_{k=1}^2 (y_{j_k} - t_{j_k})^2}} \times \sum_{k=1}^2 2(y_j - t_j)_k \\ &= \frac{\sum_{k=1}^2 2(y_j - t_j)_k}{2\sqrt{\sum_{k=1}^2 (y_{j_k} - t_{j_k})^2}} \\ &= \frac{\sum_{k=1}^2 (y_j - t_j)_k}{\sqrt{\sum_{k=1}^2 (y_{j_k} - t_{j_k})^2}} \\ &= \frac{\sum_{k=1}^2 (y_j - t_j)_k}{\|y_j - t_j\|} \end{aligned} \tag{9.1}$$

Check point 3: the constant multiple rule states that for any function f and any constant c ,

$$\frac{d}{dx} c f(x) = c \frac{d}{dx} f(x) \tag{9.2}$$

Therefore, if the minimum distance between a point y_j and set of points, T_H , is defined as:

$$s(y_j, T_H) = \min_{t_j \in T_H} \|y_j - t_j\|$$

then,

$$\frac{\partial}{\partial Y_H} \left(\frac{1}{|Y_H|} \sum_{y_j \in Y_H} s(y_j, T_H) \right) = \frac{1}{|Y_H|} \sum_{y_j \in Y_H} \min_{t_j \in T_H} \left(\frac{\sum_{k=1}^2 (y_j - t_j)_k}{\|y_j - t_j\|} \right) \quad (9.3)$$

B) Let us express the Euclidean norm between point point t_j and y_j in 2D space as:

$$\|t_j - y_j\| = \sqrt{\sum_{k=1}^2 (t_{j_k} - y_{j_k})^2} \quad (9.4)$$

Let us consider the partial derivative of $\|t_j - y_j\|$ with respect to y_j . Thus,

$$\begin{aligned} \frac{\partial}{\partial y_j} \left(\sqrt{\sum_{k=1}^2 (t_{j_k} - y_{j_k})^2} \right) &= \frac{1}{2} \times \frac{1}{\sqrt{\sum_{k=1}^2 (t_{j_k} - y_{j_k})^2}} \times \sum_{k=1}^2 -2(t_j - y_j)_k \\ &= \frac{\sum_{k=1}^2 -2(t_j - y_j)_k}{2\sqrt{\sum_{k=1}^2 (t_{j_k} - y_{j_k})^2}} \\ &= \frac{\sum_{k=1}^2 -(t_j - y_j)_k}{\sqrt{\sum_{k=1}^2 (t_{j_k} - y_{j_k})^2}} \\ &= \frac{\sum_{k=1}^2 (y_j - t_j)_k}{\|t_j - y_j\|} \end{aligned} \quad (9.5)$$

Therefore, if the minimum distance between a point t_j and set of points, Y_H is defined as:

$$s(t_j, Y_H) = \min_{y_j \in Y_H} \|t_j - y_j\| \quad (9.6)$$

then,

$$\frac{\partial}{\partial Y_H} \left(\frac{1}{|T_H|} \sum_{t_j \in T_H} s(t_j, Y_H) \right) = \frac{1}{|T_H|} \sum_{t_j \in T_H} \min_{y_j \in Y_H} \left(\frac{\sum_{k=1}^2 (y_j - t_j)_k}{\|t_j - y_j\|} \right) \quad (9.7)$$

References

- [1] Farag A, Lu L, Roth HR, Liu J, Turkbey E, Summers RM. A bottom-up approach for pancreas segmentation using cascaded superpixels and (deep) image patch labeling. *IEEE Transactions on Image Processing*. 2016;26(1):386-99. Available from: <https://doi.org/10.1109/TIP.2016.2624198>.
- [2] Yuan J, Bae E, Tai XC. A study on continuous max-flow and min-cut approaches. In: *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE; 2010. p.2217-2224.
- [3] Wu C, Tai XC. Augmented Lagrangian method, dual methods, and split Bregman iteration for ROF, vectorial TV, and high order models. *SIAM Journal on Imaging Sciences*. 2010;3(3):300-39. Available from: <https://doi.org/10.1137/090767558>.
- [4] Dollár P, Zitnick CL. Structured forests for fast edge detection. In *Proceedings of the IEEE International Conference on Computer Vision*. IEEE; 2013. p.1841-1848.
- [5] Dubuisson MP, Jain AK. A modified Hausdorff distance for object matching. In: *Proceedings of the 12th IEEE international conference on pattern recognition*. IEEE; 1994. p.566-568.
- [6] Dosselmann R, Yang XD. A comprehensive assessment of the structural similarity index. *Signal, Image and Video Processing*. 2011;5(1):81-91. Available from: <https://doi.org/10.1007/s11760-009-0144-1>.
- [7] Wang Z, Lu L, Bovik AC. Video quality assessment based on structural distortion measurement. *Signal processing: Image communication*. 2004;19(2):121-32. Available from: [https://doi.org/10.1016/S0923-5965\(03\)00076-6](https://doi.org/10.1016/S0923-5965(03)00076-6).
- [8] Henrikson J. Completeness and total boundedness of the Hausdorff metric. *MIT Undergraduate Journal of Mathematics*. 1999;1:69-80.
- [9] Hagen-Ansert S. *Textbook of Diagnostic Sonography*. 8th ed. Mosby Inc.; 2017.
- [10] Cruickshank A, Benbow E. *Pathology of the Pancreas*. 2nd ed. London: Springer London; 1995.

- [11] Williams DJ, Shah M. A fast algorithm for active contours and curvature estimation. *CVGIP: Image understanding*. 1992;55(1):14-26. Available from: [https://doi.org/10.1016/1049-9660\(92\)90003-L](https://doi.org/10.1016/1049-9660(92)90003-L).
- [12] Burute N, Nisenbaum R, Jenkins DJ, Mirrahimi A, Anthwal S, Colak E, Kirpalani A. Pancreas volume measurement in patients with Type 2 diabetes using magnetic resonance imaging-based planimetry. *Pancreatology*. 2014;14(4):268-74. Available from: <https://doi.org/10.1016/j.pan.2014.04.031>.
- [13] Asaturyan H, Villarini B. Hierarchical framework for automatic pancreas segmentation in MRI using continuous max-flow and min-cuts approach. In: *International Conference Image Analysis and Recognition 2018*. Springer, Cham; 2018. p.562-570.
- [14] Roth HR, Lu L, Farag A, Shin HC, Liu J, Turkbey EB, Summers RM. Deeporgan: Multi-level deep convolutional networks for automated pancreas segmentation. In: *International conference on medical image computing and computer-assisted intervention*. Springer, Cham; 2015. p.556-564.
- [15] Omeri AK, Matsumoto S, Kiyonaga M, Takaji R, Yamada Y, Kosen K, Mori H, Miyake H. Contour variations of the body and tail of the pancreas: evaluation with MDCT. *Japanese Journal of Radiology*. 2017;35(6):310-8. Available from: <https://doi.org/10.1007/s11604-017-0635-x>.
- [16] Villarini B, Asaturyan H, Thomas EL, Mould R, Bell JD. A framework for morphological feature extraction of organs from MR images for detection and classification of abnormalities. In: *2017 IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS)*. IEEE; 2017. p.666-671.
- [17] Cnossen WR, Drenth JP. Polycystic liver disease: an overview of pathogenesis, clinical manifestations and management. *Orphanet journal of rare diseases*. 2014;9(1):1-3. Available from: <https://doi.org/10.1186/1750-1172-9-69>.
- [18] Geraghty EM, Boone JM, McGahan JP, Jain K. Normal organ volume assessment from abdominal CT. *Abdominal imaging*. 2004;29(4):482-90. Available from: <https://doi.org/10.1007/s00261-003-0139-2>.
- [19] Chen CM, Chou YH, Tagawa N, Do Y. Computer-aided detection and diagnosis in medical imaging. *Computational and Mathematical Methods in Medicine*. 2013;2013:556-564. Available from: <https://doi.org/10.1155/2013/790608>.
- [20] Kayalibay B, Jensen G, van der Smagt P. CNN-based segmentation of medical imaging data. *ArXiv [Preprint]* 2017. Available from: <https://arxiv.org/abs/1701.03056>.

- [21] Wolz R, Chu C, Misawa K, Fujiwara M, Mori K, Rueckert D. Automated abdominal multi-organ segmentation with subject-specific atlas generation. *IEEE transactions on medical imaging*. 2013;32(9):1723-30. Available from: <https://doi.org/10.1109/TMI.2013.2265805>.
- [22] Cuingnet R, Prevost R, Lesage D, Cohen LD, Mory B, Ardon R. Automatic detection and segmentation of kidneys in 3D CT images using random forests. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Berlin, Heidelberg: Springer; 2012. p.66-74.
- [23] Tong T, Wolz R, Wang Z, Gao Q, Misawa K, Fujiwara M, Mori K, Hajnal JV, Rueckert D. Discriminative dictionary learning for abdominal multi-organ segmentation. *Medical image analysis*. 2015;23(1):92-104. Available from: <https://doi.org/10.1016/j.media.2015.04.015>.
- [24] Cai J, Lu L, Xie Y, Xing F, Yang L. Pancreas segmentation in MRI using graph-based decision fusion on convolutional neural networks. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, Cham; 2017. p.674-682.
- [25] Chu C, Oda M, Kitasaka T, Misawa K, Fujiwara M, Hayashi Y, Nimura Y, Rueckert D, Mori K. Multi-organ segmentation based on spatially-divided probabilistic atlas from 3D abdominal CT images. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Berlin, Heidelberg: Springer; 2013. p.165-172.
- [26] Campbell-Thompson ML, Filipp SL, Grajo JR, Nambam B, Beegle R, Middlebrooks EH, Gurka MJ, Atkinson MA, Schatz DA, Haller MJ. Relative pancreas volume is reduced in first-degree relatives of patients with type 1 diabetes. *Diabetes Care*. 2019;42(2):281-7. Available from: <https://doi.org/10.2337/dc18-1512>.
- [27] Wang Z, Bhatia KK, Glocker B, Marvao A, Dawes T, Misawa K, Mori K, Rueckert D. Geodesic patch-based segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, Cham; 2014. p.666-673.
- [28] Okada T, Linguraru MG, Hori M, Summers RM, Tomiyama N, Sato Y. Abdominal multi-organ segmentation from CT images using conditional shape–location and unsupervised intensity priors. *Medical image analysis*. 2015;26(1):1-8. Available from: <https://doi.org/10.1016/j.media.2015.06.009>.

- [29] Wolz R, Chu C, Misawa K, Mori K, Rueckert D. Multi-organ abdominal CT segmentation using hierarchically weighted subject-specific atlases. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Berlin, Heidelberg: Springer; 2012. p.10-17.
- [30] Cai J, Lu L, Xie Y, Xing F, Yang L. Improving deep pancreas segmentation in CT and MRI images via recurrent neural contextual learning and direct loss function. *ArXiv [Preprint]* 2017. Available from: <https://arxiv.org/abs/1707.04912>.
- [31] Milletari F, Navab N, Ahmadi SA. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: *2016 fourth international conference on 3D vision (3DV)*. IEEE; 2016. p.565-571.
- [32] Shimizu A, Kimoto T, Kobatake H, Nawano S, Shinozaki K. Automated pancreas segmentation from three-dimensional contrast-enhanced computed tomography. *International journal of computer assisted radiology and surgery*. 2010;5(1):85. Available from: <https://doi.org/10.1007/s11548-009-0384-0>.
- [33] Okada T, Linguraru MG, Yoshida Y, Hori M, Summers RM, Chen YW, Tomiyama N, Sato Y. Abdominal multi-organ segmentation of CT images based on hierarchical spatial modeling of organ interrelations. In: *International MICCAI Workshop on Computational and Clinical Challenges in Abdominal Imaging*. Berlin, Heidelberg: Springer; 2011. p.173-180.
- [34] Yushkevich PA, Wang H, Pluta J, Avants BB. From label fusion to correspondence fusion: a new approach to unbiased groupwise registration. In: *2012 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE; 2012. p. 956-963.
- [35] Zhang D, Wu G, Jia H, Shen D. Confidence-guided sequential label fusion for multi-atlas based segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Berlin, Heidelberg: Springer; 2011. p.643-650.
- [36] Dryden I, Mardia K. *Statistical shape analysis with application in R*. Chichester: Wiley; 2012.
- [37] LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521(7553):436-44. Available from: <https://doi.org/10.1038/nature14539>.
- [38] Who.int. *Diabetes*. Available from: <https://www.who.int/news-room/fact-sheets/detail/diabetes> [Accessed 7 January 2019].
- [39] Gourtsoyiannis N, Aschoff P. *Clinical MRI of the abdomen: why, how, when*. Springer Science & Business Media; 2011.

- [40] Hamm B, Forstner R. *MRI and CT of the Female Pelvis*. New York: Springer-Verlag; 2007.
- [41] McCollough CH, Schueler BA, Atwell TD, Braun NN, Regner DM, Brown DL, LeRoy AJ. Radiation exposure and pregnancy: when should we be concerned?. *Radiographics*. 2007;27(4):909-17. Available from: <https://doi.org/10.1148/rg.274065149>.
- [42] Neijenhuis MK, Kievit W, Verheesen SM, D'Agnolo HM, Gevers TJ, Drenth JP. Impact of liver volume on polycystic liver disease-related symptoms and quality of life. *United European gastroenterology journal*. 2018;6(1):81-8. Available from: <https://doi.org/10.1177/2050640617705577>.
- [43] Kistler AD, Poster D, Krauer F, Weishaupt D, Raina S, Senn O, Binet I, Spanaus K, Wüthrich RP, Serra AL. Increases in kidney volume in autosomal dominant polycystic kidney disease can be detected within 6 months. *Kidney international*. 2009;75(2):235-41. Available from: <https://doi.org/10.1038/ki.2008.558>.
- [44] Macauley M, Percival K, Thelwall PE, Hollingsworth KG, Taylor R. Altered volume, morphology and composition of the pancreas in type 2 diabetes. *PloS one*. 2015;10(5):e0126825. Available from: <https://doi.org/10.1371/journal.pone.0126825>.
- [45] Al-Mrabeh A, Hollingsworth KG, Shaw JA, McConnachie A, Sattar N, Lean ME, Taylor R. 2-year remission of type 2 diabetes and pancreas morphology: a post-hoc analysis of the DiRECT open-label, cluster-randomised trial. *The Lancet Diabetes & Endocrinology*. 2020;8(12):939-48. Available from: [https://doi.org/10.1016/S2213-8587\(20\)30303-X](https://doi.org/10.1016/S2213-8587(20)30303-X).
- [46] Shintaku M, Arimoto A, Sakita N. Serous cystadenocarcinoma of the pancreas. *Pathology international*. 2005;55(7):436-9. Available from: <https://doi.org/10.1111/j.1440-1827.2005.01850.x>.
- [47] Beger H, Peng S, Sarr M, Nakao A. *Pancreatic Cancer, Cystic and Endocrine Neoplasms*. John Wiley & Sons; 2015.
- [48] Goh BK, Thng CH, Tan DM, Low AS, Wong JS, Cheow PC, Chow PK, Chung AY, Wong WK, Ooi LL. Evaluation of the Sendai and 2012 International Consensus Guidelines based on cross-sectional imaging findings performed for the initial triage of mucinous cystic lesions of the pancreas: a single institution experience with 114 surgically treated patients. *The American Journal of Surgery*. 2014;208(2):202-9. Available from: <https://doi.org/10.1016/j.amjsurg.2013.09.031>.

- [49] Tanaka M, Chari S, Adsay V, Castillo FD, Falconi M, Shimizu M, Yamaguchi K, Yamao K, Matsuno S. International consensus guidelines for management of intraductal papillary mucinous neoplasms and mucinous cystic neoplasms of the pancreas. *Pancreatology*. 2006;6(1-2):17-32. Available from: <https://doi.org/10.1159/000090023>.
- [50] Brugge WR, Lauwers GY, Sahani D, Fernandez-del Castillo C, Warshaw AL. Cystic neoplasms of the pancreas. *New England Journal of Medicine*. 2004;351(12):1218-26. Available from: <https://doi.org/10.1056/NEJMra031623>.
- [51] Curry CA, Eng J, Horton KM, Urban B, Siegelman S, Kuszyk BS, Fishman EK. CT of primary cystic pancreatic neoplasms: can CT be used for patient triage and treatment?. *American Journal of Roentgenology*. 2000;175(1):99-103. Available from: <https://doi.org/10.2214/ajr.175.1.1750099>.
- [52] Banks PA, Bollen TL, Dervenis C, Gooszen HG, Johnson CD, Sarr MG, Tsiotos GG, Vege SS. Classification of acute pancreatitis—2012: revision of the Atlanta classification and definitions by international consensus. *Gut*. 2013;62(1):102-11. Available from: <http://dx.doi.org/10.1136/gutjnl-2012-302779>.
- [53] Piironen A. Severe acute pancreatitis: contrast-enhanced CT and MRI features. *Abdominal Radiology*. 2001;26(3):225. Available from: <https://doi.org/10.1007/s002610000153>.
- [54] Arvanitakis M, Delhayé M, De Maertelaere V, Bali M, Winant C, Cop-pens E, Jeanmart J, Zalcmán M, Van Gansbeke D, Devière J, Matos C. Computed tomography and magnetic resonance imaging in the assessment of acute pancreatitis. *Gastroenterology*. 2004;126(3):715-23. Available from: <https://doi.org/10.1053/j.gastro.2003.12.006>.
- [55] Amano Y, Oishi T, Takahashi M, Kumazaki T. Nonenhanced magnetic resonance imaging of mild acute pancreatitis. *Abdominal imaging*. 2001;26(1):59-63. Available from: <https://doi.org/10.1107/s002610000104>.
- [56] Siddiqi AJ, Miller F. Chronic pancreatitis: ultrasound, computed tomography, and magnetic resonance imaging features. In: *Seminars in Ultrasound, CT and MRI* 2007. WB Saunders; 2007. p.384-394.
- [57] Tirkes T, Sandrasegaran K, Sanyal R, Sherman S, Schmidt CM, Cotte GA, Akisik F. Secretin-enhanced MR cholangiopancreatography: spectrum of findings. *Radiographics*. 2013;33(7):1889-906. Available from: <https://doi.org/10.1148/rg.337125014>.

- [58] Bolan PJ, Fink AS. Endoscopic retrograde cholangiopancreatography in chronic pancreatitis. *World journal of surgery*. 2003;27(11):1183-91. Available from: <https://doi.org/10.1007/s00268-003-7236-9>.
- [59] Fukukura Y, Fujiyoshi F, Sasaki M, Nakajo M. Pancreatic duct: morphologic evaluation with MR cholangiopancreatography after secretin stimulation. *Radiology*. 2002;222(3):674-80. Available from: <https://doi.org/10.1148/radiol.2223010684>.
- [60] Van Harten B, de Leeuw FE, Weinstein HC, Scheltens P, Biessels GJ. Brain imaging in patients with diabetes: a systematic review. *Diabetes care*. 2006;29(11):2539-48. Available from: <https://doi.org/10.2337/dc06-1637>.
- [61] Gilbeau JP, Poncelet V, Libon E, Derue G, Heller FR. The density, contour, and thickness of the pancreas in diabetics: CT findings in 57 patients. *AJR. American journal of roentgenology*. 1992;159(3):527-31. Available from: <https://doi.org/10.2214/ajr.159.3.1503017>.
- [62] American Diabetes Association. Diagnosis and classification of diabetes mellitus. *Diabetes care*. 2014;37(Supplement 1):S81-90. Available from: <https://doi.org/10.2337/dc14-S081>.
- [63] American Diabetes Association. Standards of medical care in diabetes—2015 abridged for primary care providers. *Clinical diabetes: a publication of the American Diabetes Association*. 2015;33(2):97. Available from: <https://doi.org/10.2337/diaclin.33.2.97>.
- [64] Williams AJ, Chau W, Callaway MP, Dayan CM. Magnetic resonance imaging: a reliable method for measuring pancreatic volume in type 1 diabetes. *Diabetic medicine*. 2007;24(1):35-40. Available from: <https://doi.org/10.1111/j.1464-5491.2007.02027.x>.
- [65] Steven S, Hollingsworth KG, Small PK, Woodcock SA, Pucci A, Aribisala B, Al-Mrabeh A, Daly AK, Batterham RL, Taylor R. Weight loss decreases excess pancreatic triacylglycerol specifically in type 2 diabetes. *Diabetes Care*. 2016;39(1):158-65. Available from: <https://doi.org/10.2337/dc15-0750>.
- [66] Bilgin M, Balci NC, Momtahn AJ, Bilgin Y, Klör HU, Rau WS. MRI and MRCP findings of the pancreas in patients with diabetes mellitus: compared analysis with pancreatic exocrine function determined by fecal elastase 1. *Journal of clinical gastroenterology*. 2009;43(2):165-70. Available from: <https://doi.org/10.1097/MCG.0b013e3181587912>.

- [67] Rawla P, Sunkara T, Gaduputi V. Epidemiology of pancreatic cancer: global trends, etiology and risk factors. *World journal of oncology*. 2019;10(1):10. Available from: <https://doi.org/10.14740/wjon1166>.
- [68] De Boer PT, Kroese DP, Mannor S, Rubinstein RY. A tutorial on the cross-entropy method. *Annals of operations research*. 2005;134(1):19-67. Available from: <https://doi.org/10.1007/s10479-005-5724-z>.
- [69] Yeghiazaryan V, Voiculescu I. An overview of current evaluation methods used in medical image segmentation. *Department of Computer Science, University of Oxford*. 2015.
- [70] Karasawa KI, Oda M, Kitasaka T, Misawa K, Fujiwara M, Chu C, Zheng G, Rueckert D, Mori K. Multi-atlas pancreas segmentation: Atlas selection based on vessel structure. *Medical image analysis*. 2017;39:18-28. Available from: <https://doi.org/10.1016/j.media.2017.03.006>.
- [71] Radiologyinfo.org. *Magnetic Resonance Imaging (MRI) - Body*. Available from: <https://www.radiologyinfo.org/en/info.cfm?pg=bodymr> [Accessed 21st March 2019].
- [72] Roth HR, Lu L, Lay N, Harrison AP, Farag A, Sohn A, Summers RM. Spatial aggregation of holistically-nested convolutional neural networks for automated pancreas localization and segmentation. *Medical image analysis*. 2018;45:94-107. Available from: <https://doi.org/10.1016/j.media.2018.01.006>.
- [73] Zhou Y, Xie L, Shen W, Fishman E, Yuille A. Pancreas segmentation in abdominal CT scan: a coarse-to-fine approach. *ArXiv [Preprint]* 2016. Available from: <https://arxiv.org/abs/1612.08230>.
- [74] Oda M, Shimizu N, Karasawa KI, Nimura Y, Kitasaka T, Misawa K, Fujiwara M, Rueckert D, Mori K. Regression forest-based atlas localization and direction specific atlas generation for pancreas segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, Cham; 2016. p.556-563.
- [75] Saito A, Nawano S, Shimizu A. Joint optimization of segmentation and shape prior from level-set-based statistical shape model, and its application to the automated segmentation of abdominal organs. *Medical image analysis*. 2016;28:46-65. Available from: <https://doi.org/10.1016/j.media.2015.11.003>.
- [76] Roth HR, Farag A, Lu L, Turkbey EB, Summers RM. Deep convolutional networks for pancreas segmentation in CT imaging. In: *Medical Imaging 2015: Image Processing*. International Society for Optics and Photonics; 2015. p.94131G.

- [77] Roth HR, Lu L, Farag A, Sohn A, Summers RM. Spatial aggregation of holistically-nested networks for automated pancreas segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, Cham; 2016. p.451-459.
- [78] Farag A, Lu L, Turkbey E, Liu J, Summers RM. A bottom-up approach for automatic pancreas segmentation in abdominal CT scans. In: *International MIC-CAI Workshop on Computational and Clinical Challenges in Abdominal Imaging*. Springer, Cham; 2014. p.103-113.
- [79] Hammon M, Cavallaro A, Erdt M, Dankerl P, Kirschner M, Drechsler K, Wesarg S, Uder M, Janka R. Model-based pancreas segmentation in portal venous phase contrast-enhanced CT images. *Journal of digital imaging*. 2013;26(6):1082-90. <https://doi.org/10.1007/s10278-013-9586-7>.
- [80] Huttenlocher DP, Klanderman GA, Rucklidge WJ. Comparing images using the Hausdorff distance. *IEEE Transactions on pattern analysis and machine intelligence*. 1993;15(9):850-63. Available from: <https://doi.org/10.1109/34.232073>.
- [81] Cancer Imaging Archive. *Pancreas-CT*. Available from: <https://wiki.cancerimagingarchive.net/display/Public/Pancreas-CT> [Accessed 14th March 2019].
- [82] ImageNet. *ImageNet*. Available from: <http://www.image-net.org/> [Accessed 14th March 2019].
- [83] Pascal Dataset. *Pascal Dataset*. Available from: <http://host.robots.ox.ac.uk/pascal/VOC/> [Accessed 14th March 2019].
- [84] Jim Kozubek. *The Limits of Big Data in Medical Research*. Available from: <https://blogs.scientificamerican.com/observations/the-limits-of-big-data-in-medical-research/> [Accessed 14th March 2019].
- [85] UK Biobank. *UK Biobank*. Available from: <http://www.ukbiobank.ac.uk/> [Accessed 14th March 2019].
- [86] National Institute of Health. *Biobank*. Available from: <https://allofus.nih.gov/about/program-components/biobank> [Accessed 14th March 2019].
- [87] Mayo Clinic. *Education and Research at Mayo Clinic*. Available from: <https://www.mayo.edu/> [Accessed 14th March 2019].
- [88] Oasis. *OASIS Brains Datasets*. Available from: <https://www.oasis-brains.org/#data> [Accessed 14th March 2019].

- [89] OpenfMRI. *OpenfMRI*. Available from: <https://legacy.openfmri.org/> [Accessed 19th March 2019].
- [90] mridata.org. *Welcome to mridata.org!* Available from: <http://mridata.org/> [Accessed 19th March 2019].
- [91] The NIH MRI Study of Normal Brain Development. *The NIH MRI Study of Normal Brain Development*. Available from: https://pediatricmri.nih.gov/nihpd/info/data_access.html [Accessed 19th March 2019].
- [92] MedPix. *The National Library of Medicine presents MedPix*. Available from: <https://medpix.nlm.nih.gov/home> [Accessed 19th March 2019].
- [93] Dalah E, Moraru I, Paulson E, Erickson B, Li XA. Variability of target and normal structure delineation using multimodality imaging for radiation therapy of pancreatic cancer. *International Journal of Radiation Oncology* Biology* Physics*. 2014;89(3):633-40. Available from: <https://doi.org/10.1016/j.ijrobp.2014.02.035>.
- [94] Poudel A, Fowler JL, Zielinski MC, Kilimnik G, Hara M. Stereological analyses of the whole human pancreas. *Scientific reports*. 2016;6(1):1-3. Available from: <https://doi.org/10.1038/srep34049>.
- [95] Versteijne E, Gurney-Champion OJ, van der Horst A, Lens E, Kolff MW, Buijsen J, Ebrahimi G, Neelis KJ, Rasch CR, Stoker J, van Herk M. Considerable interobserver variation in delineation of pancreatic cancer on 3DCT and 4DCT: a multi-institutional study. *Radiation Oncology*. 2017;12(1):1-0. Available from: <https://doi.org/10.1186/s13014-017-0777-0>.
- [96] Ross BA, Jeffrey Jr RB, Mindelzun RE. Normal variations in the lateral contour of the head and neck of the pancreas mimicking neoplasm: evaluation with dual-phase helical CT. *AJR. American journal of roentgenology*. 1996;166(4):799-801. Available from: <https://doi.org/10.2214/ajr.166.4.8610553>.
- [97] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *ArXiv [Preprint]* 2014. Available from: <https://arxiv.org/abs/1409.1556>.
- [98] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *ArXiv [Preprint]* 2015. Available from: <https://arxiv.org/abs/1502.03167>.

- [99] Lin G, Shen C, Van Den Hengel A, Reid I. Efficient piecewise training of deep structured models for semantic segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Computer Vision Foundation; 2016. p.3194-3203.
- [100] Badrinarayanan V, Kendall A, Cipolla R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*. 2017;39(12):2481-95. Available from: <https://doi.org/10.1109/TPAMI.2016.2644615>.
- [101] PyImageSearch. *Softmax Classifiers Explained - PyImageSearch*. Available from: <https://www.pyimagesearch.com/2016/09/12/softmax-classifiers-explained/> [Accessed 20th February 2019].
- [102] Jia Y, Shelhamer E, Donahue J, Karayev S, Long J, Girshick R, Guadarrama S, Darrell T. Caffe: Convolutional architecture for fast feature embedding. In: *Proceedings of the 22nd ACM international conference on Multimedia*. 2014. p.675-678.
- [103] Eigen D, Fergus R. Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture. In: *Proceedings of the IEEE International Conference on Computer Vision*. Computer Vision Foundation; 2015. p.2650-2658.
- [104] Achanta R, Shaji A, Smith K, Lucchi A, Fua P, Süsstrunk S. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE transactions on pattern analysis and machine intelligence*. 2012;34(11):2274-82. Available from: <https://doi.org/10.1109/TPAMI.2012.120>.
- [105] Gou S, Wu J, Liu F, Lee P, Rapacchi S, Hu P, Sheng K. Feasibility of automated pancreas segmentation based on dynamic MRI. *The British journal of radiology*. 2014;87(1044):20140248. Available from: <https://doi.org/10.1259/bjr.20140248>.
- [106] Suzuki T, Takizawa H, Kudo H, Okada T. Interactive segmentation of pancreases from abdominal CT images by use of the graph cut technique with probabilistic atlases. In: *Innovation in Medicine and Healthcare 2015*. Springer, Cham; 2016. p.575-584.
- [107] Perez L, Wang J. The effectiveness of data augmentation in image classification using deep learning. *ArXiv [Preprint]* 2017. Available from: <https://arxiv.org/abs/1712.04621>.
- [108] Wong, S.C., Gatt, A., Stamatescu, V. and McDonnell, M.D. (2016). Understanding data augmentation for classification: when to warp?. *arXiv preprint arXiv:1609.08764*.

- [109] Yaeger LS, Lyon RF, Webb BJ. Effective training of a neural network character classifier for word recognition. In: *Advances in neural information processing systems*. 1997. p.807-816.
- [110] Heerkens HD, Hall WA, Li XA, Knechtges P, Dalah E, Paulson ES, van den Berg CA, Meijer GJ, Koay EJ, Crane CH, Aitken K. Recommendations for MRI-based contouring of gross tumor volume and organs at risk for radiation therapy of pancreatic cancer. *Practical radiation oncology*. 2017;7(2):126-36. Available from: <https://doi.org/10.1016/j.prro.2016.10.006>.
- [111] Andersen V, Sonne J, Sletting S, Prip A. The volume of the liver in patients correlates to body weight and alcohol consumption. *Alcohol and Alcoholism*. 2000;35(5):531-2. Available from: <https://doi.org/10.1093/alcalc/35.5.531>.
- [112] Blomquist L, Wang Y, Kimiaei S, Jacobsson H. Change in size, shape and radiocolloid uptake of the alcoholic liver during alcohol withdrawal, as demonstrated by single photon emission computed tomography. *Journal of hepatology*. 1994;21(3):417-23. Available from: [https://doi.org/10.1016/S0168-8278\(05\)80322-8](https://doi.org/10.1016/S0168-8278(05)80322-8).
- [113] Sato H, Takase S, Takada A. Changes in liver and spleen volumes in alcoholic liver disease. *Journal of hepatology*. 1989;8(2):150-7. Available from: [https://doi.org/10.1016/0168-8278\(89\)90002-0](https://doi.org/10.1016/0168-8278(89)90002-0).
- [114] Urata K, Kawasaki S, Matsunami H, Hashikura Y, Ikegami T, Ishizone S, Momose Y, Komiyama A, Makuuchi M. Calculation of child and adult standard liver volume for liver transplantation. *Hepatology*. 1995;21(5):1317-21. Available from: <https://doi.org/10.1002/hep.1840210515>.
- [115] Herold C, Heinz R, Niedobitek G, Schneider T, Hahn EG, Schuppan D. Quantitative testing of liver function in relation to fibrosis in patients with chronic hepatitis B and C. *Liver*. 2001;21(4):260-5. Available from: <https://doi.org/10.1034/j.1600-0676.2001.021004260.x>.
- [116] Chapman AB, Bost JE, Torres VE, Guay-Woodford L, Bae KT, Landsittel D, Li J, King BF, Martin D, Wetzel LH, Lockhart ME. Kidney volume and functional outcomes in autosomal dominant polycystic kidney disease. *Clinical Journal of the American Society of Nephrology*. 2012;7(3):479-86. Available from: <https://doi.org/10.2215/CJN.09500911>.
- [117] Tangri N, Hougen I, Alam A, Perrone R, McFarlane P, Pei Y. Total kidney volume as a biomarker of disease progression in autosomal dominant polycystic kidney disease. *Canadian journal of kidney health and disease*. 2017;4:2054358117693355. Available from: <https://doi.org/10.1177/2054358117693355>.

- [118] Gale HI, Bobbitt CA, Setty BN, Sprinz PG, Doros G, Williams DD, Morrison TC, Kalajian TA, Tu P, Mundluru SN, Castro-Aragon I. Expected sonographic appearance of the spleen in children and young adults with sickle cell disease: an update. *Journal of Ultrasound in Medicine*. 2016;35(8):1735-45. Available from: <https://doi.org/10.7863/ultra.15.09023>.
- [119] Ahmed AM. *Evaluation of Visceral Leishmaniasis in Gadarif State Using Ultrasonography* [Doctoral dissertation]. Sudan University of Science and Technology; 2017.
- [120] Yildiz AE, Ariyurek MO, Karcaaltincaba M. Splenic anomalies of shape, size, and location: pictorial essay. *The Scientific World Journal*. 2013;2013. Available from: <https://doi.org/10.1155/2013/321810>.
- [121] Silagy C, Shelby-James T, Sage M, Wallage A. Patient-detected diurnal changes in spleen volume. *The Lancet*. 1998;352(9129):710. Available from: [https://doi.org/10.1016/S0140-6736\(05\)60828-8](https://doi.org/10.1016/S0140-6736(05)60828-8).
- [122] Zargar H, Almassi N, Kovac E, Ercole C, Remer E, Rini B, Stephenson A, Garcia JA, Grivas P. Change in psoas muscle volume as a predictor of outcomes in patients treated with chemotherapy and radical cystectomy for muscle-invasive bladder cancer. *Bladder Cancer*. 2017;3(1):57-63. Available from: <https://doi.org/10.3233/BLC-160080>.
- [123] Yoshikawa T, Takano M, Miyamoto M, Yajima I, Shimizu Y, Aizawa Y, Suguchi Y, Moriiwa M, Aoyama T, Soyama H, Goto T. Psoas muscle volume as a predictor of peripheral neurotoxicity induced by primary chemotherapy in ovarian cancers. *Cancer Chemotherapy and Pharmacology*. 2017;80(3):555-61. Available from: <https://doi.org/10.1007/s00280-017-3395-5>.
- [124] Erdt M, Kirschner M, Drechsler K, Wesarg S, Hammon M, Cavallaro A. Automatic pancreas segmentation in contrast enhanced CT data using learned spatial anatomy and texture descriptors. In: *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. IEEE; 2011. p.2076-2082.
- [125] Qi CR, Su H, Nießner M, Dai A, Yan M, Guibas LJ. Volumetric and Multi-View CNNs for Object Classification on 3D Data. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. IEEE; 2016. p.5648-5656.
- [126] Roth HR, Lu L, Liu J, Yao J, Seff A, Cherry K, Kim L, Summers RM. Improving computer-aided detection using convolutional neural networks and random view aggregation. *IEEE transactions on medical imaging*. 2015;35(5):1170-81. Available from: <https://doi.org/10.1109/TMI.2015.2482920>.

- [127] Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, Cham; 2016. p.424-432.
- [128] Bui TD, Shin J, Moon T. 3D Densely Convolutional Networks for Volumetric Segmentation. *ArXiv [Preprint]* 2017. Available from: <https://arxiv.org/abs/1709.03199>.
- [129] Xiang D, Bagci U, Jin C, Shi F, Zhu W, Yao J, Sonka M, Chen X. CorteXpert: A model-based method for automatic renal cortex segmentation. *Medical image analysis*. 2017;42:257-73. Available from: <https://doi.org/10.1016/j.media.2017.06.010>.
- [130] Cai J, Lu L, Xing F, Yang L. Pancreas segmentation in CT and MRI images via domain specific network designing and recurrent neural contextual learning. *ArXiv [Preprint]* 2018. Available from: <https://arxiv.org/abs/1803.11303>.
- [131] Xia Y, Xie L, Liu F, Zhu Z, Fishman EK, Yuille AL. Bridging the gap between 2d and 3d organ segmentation with volumetric fusion net. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, Cham; 2018. p.445-453.
- [132] Roth HR, Lu L, Lay N, Harrison AP, Farag A, Sohn A, Summers RM. Spatial aggregation of holistically-nested convolutional neural networks for automated pancreas localization and segmentation. *Medical image analysis*. 2018;45:94-107. Available from: <https://doi.org/10.1016/j.media.2018.01.006>.
- [133] Gibson E, Giganti F, Hu Y, Bonmati E, Bandula S, Gurusamy K, Davidson B, Pereira SP, Clarkson MJ, Barratt DC. Automatic multi-organ segmentation on abdominal CT with dense v-networks. *IEEE transactions on medical imaging*. 2018;37(8):1822-34. Available from: <https://doi.org/10.1109/TMI.2018.2806309>.
- [134] Liu F, Xie L, Xia Y, Fishman E, Yuille A. Joint shape representation and classification for detecting PDAC. In: *International Workshop on Machine Learning in Medical Imaging*. Springer, Cham; 2019. p.212-220.
- [135] Brock A, Lim T, Ritchie JM, Weston N. Generative and discriminative voxel modeling with convolutional neural networks. *ArXiv [Preprint]* 2016. Available from: <https://arxiv.org/abs/1608.04236>.
- [136] Valindria VV, Lavdas I, Cerrolaza J, Aboagye EO, Rockall AG, Rueckert D, Glocker B. Small Organ Segmentation in Whole-Body MRI Using a Two-Stage

FCN and Weighting Schemes. In: *Machine Learning in Medical Imaging: 9th International Workshop, MLMI 2018, Held in Conjunction with MICCAI 2018, Granada, Spain*. Springer; 2018. p.346.

- [137] Larsson M, Zhang Y, Kahl F. Robust abdominal organ segmentation using regional convolutional neural networks. *Applied Soft Computing*. 2018;70:465-71. Available from: <https://doi.org/10.1016/j.asoc.2018.05.038>.
- [138] Lavdas I, Glocker B, Kamnitsas K, Rueckert D, Mair H, Sandhu A, Taylor SA, Aboagye EO, Rockall AG. Fully automatic, multiorgan segmentation in normal whole body magnetic resonance imaging (MRI), using classification forests (CFs), convolutional neural networks (CNNs), and a multi-atlas (MA) approach. *Medical physics*. 2017;44(10):5210-20. Available from: <https://doi.org/10.1002/mp.12492>.
- [139] Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions. *ArXiv [Preprint]* 2015. Available from: <https://arxiv.org/abs/1511.07122>.
- [140] Zhou Y, Xie L, Shen W, Wang Y, Fishman EK, Yuille AL. A fixed-point model for pancreas segmentation in abdominal CT scans. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, Cham; 2017. p.693-701.
- [141] Wang Y, Zhou Y, Tang P, Shen W, Fishman EK, Yuille AL. Training multi-organ segmentation networks with sample selection by relaxed upper confident bound. In: *International conference on medical image computing and computer-assisted intervention*. Springer, Cham; 2018. p.434-442.
- [142] Kamiya N, Zhou X, Chen H, Muramatsu C, Hara T, Yokoyama R, Kanematsu M, Hoshi H, Fujita H. Automated segmentation of psoas major muscle in X-ray CT images by use of a shape model: preliminary study. *Radiological physics and technology*. 2012;5(1):5-14. Available from: <https://doi.org/10.1007/s12194-011-0127-0>.
- [143] MathWorks. *MathWorks*. Available from: <https://uk.mathworks.com/> [Accessed 2nd January 2019].
- [144] MathWorks. *Image Processing Toolbox*. Available from: <https://uk.mathworks.com/products/image.html> [Accessed 2nd January 2019].
- [145] NiftyNet. *NiftyNet*. Available from: <http://www.niftynet.io/> [Accessed 2nd January 2019].
- [146] TensorFlow. *TensorFlow*. Available from: <https://www.tensorflow.org/> [Accessed 2nd January 2019].

- [147] Ubuntu. *Ubuntu*. Available from: <https://www.ubuntu.com/> [Accessed 2nd January 2019].
- [148] Colaboratory. *Welcome to Colaboratory!*. Available from: <https://colab.research.google.com/notebooks/welcome.ipynb>. [Accessed 2nd January 2019].
- [149] Torch. *Torch*. Available from: <http://torch.ch/> [Accessed 2nd January 2019].
- [150] PyTorch. *PyTorch*. Available from: <https://pytorch.org/> [Accessed 2nd January 2019].
- [151] Numpy. *Numpy*. Available from: <http://www.numpy.org/> [Accessed 2nd January 2019].
- [152] Caffe. *Caffe*. Available from: <http://caffe.berkeleyvision.org/> [Accessed 2nd January 2019].
- [153] Keras. *Keras*. Available from: <https://keras.io/> [Accessed 2nd January 2019].
- [154] Liu Y, Kot A, Drakopoulos F, Fedorov A, Enquobahrie A, Clatz O, Chrisochoides N. An ITK implementation of physics-based non-rigid registration method. *Insight J*. 2012. Available from: <https://www.insight-journal.org/browse/publication/876>.
- [155] Johnson HJ, McCormick MM, Ibáñez L. *The ITK Software Guide Book 1: Introduction and Development Guidelines-Volume 1*. Kitware, Inc.; 2015.
- [156] Penny WD, Friston KJ, Ashburner JT, Kiebel SJ, Nichols TE, (eds.). *Statistical parametric mapping: the analysis of functional brain images*. Elsevier; 2011.
- [157] Larrabide I, Omedas P, Martelli Y, Planes X, Nieber M, Moya JA, Butakoff C, Sebastián R, Camara O, De Craene M, Bijmens BH. GIMIAS: an open source framework for efficient development of research tools and clinical prototypes. In: *International Conference on Functional Imaging and Modeling of the Heart*. Berlin, Heidelberg: Springer; 2009. p.417-426.
- [158] Peter Junaidy. *Liver, pancreas, gallbladder, duodenum and bile passage labeled. Scientifically accurate*. Available from: <https://www.dreamstime.com/royalty-free-stock-images-liver-pancreas-gallbladder-duodenum-bile-pa-image24832999/> [Accessed 3rd March 2020].
- [159] Yushkevich PA, Piven J, Hazlett HC, Smith RG, Ho S, Gee JC, Gerig G. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage*. 2006;31(3):1116-28.

- [160] Mengler, L., Khmelinskii, A., Diedenhofen, M., Po, C., Staring, M., Lelieveldt, B.P. and Hoehn, M. (2014). Brain maturation of the adolescent rat cortex and striatum: changes in volume and myelination. *Neuroimage*, 84, pp.35-44. Available from: <https://doi.org/10.1016/j.neuroimage.2006.01.015>.
- [161] Welcome to PyTorch Tutorials. *Welcome to PyTorch Tutorials*. Available from: <https://pytorch.org/tutorials/> [Accessed 3rd January 2019].
- [162] Get Started with TensorFlow. *Get Started with TensorFlow*. Available from: <https://www.tensorflow.org/tutorials/> [Accessed 3rd January 2019].
- [163] VTK. *VTK*. Available from: <https://vtk.org> [Accessed 3rd January 2019].
- [164] Agrawal S, Agrawal J. A Study on Open Source Software in Health Care System. *REMOTE SENSING AND GIS*. 2012. Available from: http://csi-india.org.in/Communications/CSIC_dec_2016.pdf#page=36.
- [165] Fagiolo G, Waldman A, Hajnal JV. A simple procedure to improve FMRIB software library brain extraction tool performance. *The British journal of radiology*. 2008;81(963):250-1. Available from: <https://doi.org/10.1259/bjr/12956156>.
- [166] Cook PA, Bai Y, Nedjati-Gilani SK, Seunarine KK, Hall MG, Parker GJ, Alexander DC. Camino: open-source diffusion-MRI reconstruction and processing. In: *14th scientific meeting of the international society for magnetic resonance in medicine 2006*. Seattle, WA, USA; 2006. p.2759.
- [167] Lee LK, Liew SC. A survey of medical image processing tools. In: *2015 4th International Conference on Software Engineering and Computer Systems (ICSECS)*. IEEE; 2015. p.171-176.
- [168] Christ PF, Ettlinger F, Grün F, Elshaera ME, Lipkova J, Schlecht S, Ahmaddy F, Tatavarty S, Bickel M, Bilic P, Rempfler M. Automatic liver and tumor segmentation of CT and MRI volumes using cascaded fully convolutional neural networks. *ArXiv [Preprint]* 2017. Available from: <https://arxiv.org/abs/1702.05970>.
- [169] Li C, Wang X, Eberl S, Fulham M, Yin Y, Chen J, Feng DD. A likelihood and local constraint level set model for liver tumor segmentation from CT volumes. *IEEE Transactions on Biomedical Engineering*. 2013;60(10):2967-77. Available from: <https://doi.org/10.1109/TBME.2013.2267212>.
- [170] Ben-Cohen A, Diamant I, Klang E, Amitai M, Greenspan H. Fully convolutional network for liver segmentation and lesions detection. In: *Deep learning and data labeling for medical applications*. Springer, Cham; 2016. p.77-85.

- [171] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical image computing and computer-assisted intervention*. Springer, Cham; 2015. p.234-241.
- [172] Mostafa A, Fouad A, Abd Elfattah M, Hassanien AE, Hefny H, Zhu SY, Schaefer G. CT liver segmentation using artificial bee colony optimisation. *Procedia Computer Science*. 2015;60:1622-30. Available from: <https://doi.org/10.1016/j.procs.2015.08.272>.
- [173] Roth HR, Oda H, Zhou X, Shimizu N, Yang Y, Hayashi Y, Oda M, Fujiwara M, Misawa K, Mori K. An application of cascaded 3D fully convolutional networks for medical image segmentation. *Computerized Medical Imaging and Graphics*. 2018;66:90-9. Available from: <https://doi.org/10.1016/j.compmedimag.2018.03.001>.
- [174] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015. p.3431-3440.
- [175] McCune B, Grace JB, Urban DL. *Analysis of ecological communities*. Gleneden Beach, OR: MjM software design; 2002.
- [176] Medicalstocks. *Blood supply of the pancreas 3d medical vector illustration on white background*. Available from: <https://www.shutterstock.com/image-vector/blood-supply-pancreas-3d-medical-vector-1309328005> [Accessed 3rd March 2020].
- [177] Feng, L., Grimm, R., Block, K.T., Chandarana, H., Kim, S., Xu, J., Axel, L., Sodickson, D.K. and Otazo, R. (2014). Golden-angle radial sparse parallel MRI: combination of compressed sensing, parallel imaging, and golden-angle radial sampling for fast and flexible dynamic volumetric MRI. *Magnetic resonance in medicine*, 72(3), pp.707-717.
- [178] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016. p.770-778.
- [179] Jégou S, Drozdal M, Vazquez D, Romero A, Bengio Y. The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation. In: *Proceedings of the IEEE conference on computer vision and pattern recognition workshop*. 2017. p.11-19.
- [180] Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017. p.4700-4708.

- [181] Heywang-Köbrunner SH, Viehweg P, Heinig A, Küchler CH. Contrast-enhanced MRI of the breast: accuracy, value, controversies, solutions. *European journal of radiology*. 1997;24(2):94-108. Available from: [https://doi.org/10.1016/S0720-048X\(96\)01142-4](https://doi.org/10.1016/S0720-048X(96)01142-4).
- [182] Westra C, Dialani V, Mehta TS, Eisenberg RL. Using T2-weighted sequences to more accurately characterize breast masses seen on MRI. *American Journal of Roentgenology*. 2014;202(3):W183-90. Available from: <https://doi.org/10.2214/AJR.13.11266>.
- [183] Baraliakos X, Hermann KA, Landewe R, Listing J, Golder W, Brandt J, Rudwaleit M, Bollow M, Sieper J, van der Heijde DM, Braun J. Assessment of acute spinal inflammation in patients with ankylosing spondylitis by magnetic resonance imaging: a comparison between contrast enhanced T1 and short tau inversion recovery (STIR) sequences. *Annals of the Rheumatic Diseases*. 2005;64(8):1141-4. Available from: <http://dx.doi.org/10.1136/ard.2004.031609>.
- [184] Abdi H, Williams LJ. Principal component analysis. *Wiley interdisciplinary reviews: computational statistics*. 2010;2(4):433-59. Available from: <https://doi.org/10.1002/wics.101>.
- [185] Tummala P, Junaidi O, Agarwal B. Imaging of pancreatic cancer: An overview. *Journal of gastrointestinal oncology*. 2011;2(3):168. Available from: <https://doi.org/10.3978/j.issn.2078-6891.2011.036>.
- [186] Hurwitz LM, Yoshizumi T, Reiman RE, Goodman PC, Paulson EK, Frush DP, Toncheva G, Nguyen G, Barnes L. Radiation dose to the fetus from body MDCT during early gestation. *American Journal of Roentgenology*. 2006;186(3):871-6. Available from: <https://doi.org/10.2214/AJR.04.1915>.
- [187] Wu Y, Ding Y, Tanaka Y, Zhang W. Risk factors contributing to type 2 diabetes and recent advances in the treatment and prevention. *International journal of medical sciences*. 2014;11(11):1185. Available from: <https://doi.org/10.7150/ijms.10001>.
- [188] University of Pittsburgh Medical Center, CC BY-SA. *A pancreatic cyst, marked by an arrow, as shown on an MRI* [Image]. Available from: <https://www.upmc.com/media/news/pancreaseq> [Accessed 7th January 2019].
- [189] Werthmann PG, Kempenich R, Lang-Avérous G, Kienle GS. Long-term survival of a patient with advanced pancreatic cancer under adjunct treatment with Viscum album extracts: A case report. *World Journal of Gastroenterology*. 2019;25(12):1524. Available from: <https://doi.org/10.3748/wjg.v25.i12.1524>.

- [190] Türkvatan A, Erden A, Türkoğlu MA, Seçil M, Yener Ö. Imaging of acute pancreatitis and its complications. Part 1: acute pancreatitis. *Diagnostic and Interventional Imaging*. 2015;96(2):151-60. Available from: <https://doi.org/10.1016/j.diii.2013.12.017>.
- [191] Miller FH, Keppke AL, Dalal K, Ly JN, Kamler VA, Sica GT. MRI of pancreatitis and its complications: part 1, acute pancreatitis. *American Journal of Roentgenology*. 2004;183(6):1637-44. Available from: <https://doi.org/10.2214/ajr.183.6.01831637>.
- [192] Xiao B, Zhang XM, Tang W, Zeng NL, Zhai ZH. Magnetic resonance imaging for local complications of acute pancreatitis: a pictorial review. *World Journal Of Gastroenterology: WJG*. 2010;16(22):2735. Available from: <https://doi.org/10.3748/wjg.v16.i22.2735>.
- [193] Xiao B, Zhang XM. Magnetic resonance imaging for acute pancreatitis. *World Journal of Radiology*. 2010;2(8):298. Available from: <https://doi.org/10.4329/wjr.v2.i8.298>.
- [194] Kamat R, Gupta P, Rana S. Imaging in chronic pancreatitis: State of the art review. *The Indian Journal of Radiology & Imaging*. 2019;29(2):201. Available from: https://doi.org/10.4103/ijri.IJRI_484_18.
- [195] Semelka RC, Ascher SM. MR imaging of the pancreas. *Radiology*. 1993;188(3):593-602. Available from: <https://doi.org/10.1148/radiology.188.3.8351317>.
- [196] Balci C. MRI assessment of chronic pancreatitis. *Diagnostic and interventional radiology*. 2011;17(3):249. Available from: <https://doi.org/10.4261/1305-3825.DIR.3889-10.0>.
- [197] Gottapu RD, Dagli CH. DenseNet for anatomical brain segmentation. *Procedia Computer Science*. 2018;140:179-85. Available from: <https://doi.org/10.1016/j.procs.2018.10.327>.
- [198] Zhang R, Zhao L, Lou W, Abrigo JM, Mok VC, Chu WC, Wang D, Shi L. Automatic segmentation of acute ischemic stroke from DWI using 3-D fully convolutional DenseNets. *IEEE transactions on medical imaging*. 2018;37(9):2149-60. Available from: <https://doi.org/10.1109/TMI.2018.2821244>.
- [199] Li Y, Xie X, Shen L, Liu S. Reverse active learning based atrous DenseNet for pathological image classification. *BMC bioinformatics*. 2019;20(1):445. Available from: <https://doi.org/10.1186/s12859-019-2979-y>.

- [200] Xu X, Lin J, Tao Y, Wang X. An Improved DenseNet Method Based on Transfer Learning for Fundus Medical Images. In: *2018 7th International Conference on Digital Home (ICDH) 2018*. IEEE; 2018. p.137-140.
- [201] Kreso I, Segvic S, Krapac J. Ladder-style densenets for semantic segmentation of large natural images. In: *Proceedings of the IEEE International Conference on Computer Vision Workshops*. 2017. p.238-245.
- [202] Wu CL, Zhang Q, Zhou J, Yang H, Li Y. Text Detection and Recognition for Natural Scene Images Using Deep Convolutional Neural Networks. *Computers, Materials & Continua*. 2019;61(1):289-300. Available from: <https://doi.org/10.32604/cmc.2019.05990>.
- [203] Kingma DP, Ba J. Adam: A method for stochastic optimization. *ArXiv [Preprint]* 2014. Available from: <https://arxiv.org/abs/1412.6980>.
- [204] Chen H, Qi Y, Yin Y, Li T, Liu X, Li X, Gong G, Wang L. MMFNet: A multi-modality MRI fusion network for segmentation of nasopharyngeal carcinoma. *Neurocomputing*. 2020. Available from: <https://doi.org/10.1016/j.neucom.2020.02.002>.
- [205] Li C, Sun H, Liu Z, Wang M, Zheng H, Wang S. Learning Cross-Modal Deep Representations for Multi-Modal MR Image Segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, Cham; 2019. p.57-65.
- [206] Rahman MA, Wang Y. Optimizing intersection-over-union in deep neural networks for image segmentation. In: *International symposium on visual computing*. Springer, Cham; 2016. p.234-244.
- [207] Sudre CH, Li W, Vercauteren T, Ourselin S, Cardoso MJ. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In: *Deep learning in medical image analysis and multimodal learning for clinical decision support*. Springer, Cham; 2017. p.240-248.
- [208] Arbelaez P, Maire M, Fowlkes C, Malik J. Contour detection and hierarchical image segmentation. *IEEE transactions on pattern analysis and machine intelligence*. 2010;33(5):898-916. Available from: <https://doi.org/10.1109/TPAMI.2010.161>.
- [209] Ogurtsova K, da Rocha Fernandes JD, Huang Y, Linnenkamp U, Guariguata L, Cho NH, Cavan D, Shaw JE, Makaroff LE. IDF Diabetes Atlas: Global estimates for the prevalence of diabetes for 2015 and 2040.

Diabetes research and clinical practice. 2017;128:40-50. Available from: <https://doi.org/10.1016/j.diabres.2017.03.024>.

- [210] Hruban RH, Goggins M, Parsons J, Kern SE. Progression model for pancreatic cancer. *Clinical cancer research*. 2000;6(8):2969-72.
- [211] Warshaw AL, Castillo CF. Pancreatic carcinoma. *New England Journal of Medicine*. 1992;326(7):455-65. Available from: <https://doi.org/10.1056/NEJM199202133260706>.
- [212] Rozenblum E, Schutte M, Goggins M, Hahn SA, Panzer S, Zahurak M, Goodman SN, Sohn TA, Hruban RH, Yeo CJ, Kern SE. Tumor-suppressive pathways in pancreatic carcinoma. *Cancer research*. 1997;57(9):1731-4.
- [213] Steinkraus D, Buck I, Simard PY. Using GPUs for machine learning algorithms. In: *Eighth International Conference on Document Analysis and Recognition (ICDAR '05)*. IEEE; 2005. p.1115-1120.
- [214] Mayo Clinic Staff. *MRI*. Available from: <https://www.mayoclinic.org/tests-procedures/mri/about/pac-20384768> [Accessed 25th November 2020].
- [215] NHS. *Who can have one-MRI scan*. Available from: <https://www.nhs.uk/conditions/mri-scan/who-can-have-it> [Accessed 25th November 2020].
- [216] RadiologyInfo.org. *Computer Tomography (CT) Safety During Pregnancy*. Available from: <https://www.radiologyinfo.org/en/info.cfm?pg=safety-ct-pregnancy> [Accessed 25th November 2020].
- [217] Mayo Clinic Staff. *CT Scan*. Available at: <https://www.mayoclinic.org/tests-procedures/ct-scan/about/pac-20393675> [Accessed 25th November 2020].
- [218] Dubey AK, Jain V. Comparative Study of Convolution Neural Network's ReLu and Leaky-ReLu Activation Functions. In: *Applications of Computing, Automation and Wireless Systems in Electrical Engineering*. Springer; 2019. p.873-880.
- [219] Wang T, Qin Z, Zhu M. An ELU network with total variation for image denoising. In: *International Conference on Neural Information Processing*. Springer, Cham; 2017. p.227-237.
- [220] Klambauer G, Unterthiner T, Mayr A, Hochreiter S. Self-normalizing neural networks. *Advances in neural information processing systems*. 2017;30:971-80.
- [221] Saxe AM, McClelland JL, Ganguli S. Exact solutions to the nonlinear dynamics of learning in deep linear neural networks. *ArXiv [Preprint]* 2013. Available from: <https://arxiv.org/abs/1312.6120>.

- [222] Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks. In: *Proceedings of the thirteenth international conference on artificial intelligence and statistics*. 2010. p.249-256.